DEVELOPMENT OF A CLINICALLY-TARGETED HUMAN ACTIVITY RECOGNITION SYSTEM TO AID THE PROSTHETIC REHABILITATION OF INDIVIDUALS WITH LOWER LIMB AMPUTATION IN FREE LIVING CONDITIONS

By

ALEXANDER JAMIESON

A dissertation submitted in partial fulfillment of
the requirements for the degree of

DOCTOR OF PHILOSOPHY

UNIVERSITY OF STRATHCLYDE
School of Biomedical Engineering

OCTOBER 2021

DECLARATION OF AUTHENTICITY

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

Alexander Jamieson

ACKNOWLEDGMENT

To my supervisors, Arjan and Laura, thank you for your guidance and unwavering support over the past three years. This thesis would not have been possible without your knowledge and resources. It has been a genuine pleasure to work with you.

Thank you also to Vladimir and Lina, who were brought into the project to provide additional guidance in regards to the machine learning aspect after I was struggling to acquire meaningful results. Without your help, I would not have completed the thesis on time.

To my sponsor company PAL Technologies, and especially Dougie, David, and Nicholas, thank you for providing the ActivPALs for my project, as well as your invaluable input and feedback in regards to the direction of my thesis.

Thank you to all my volunteers who have taken part in my projects over the past three years. I have been extremely lucky to have your input, especially during these challenging times.

Thank you to all my friends and family for your moral support during my time working on my thesis. Thanks to you all, I have seen it through to the end.

Finally I'd like to give a general thanks to the University of Strathclyde and its staff for providing the right amount of support and resources whenever I have needed it.

Abstract

Aim: Healthcare Professionals (HCPs) that specialize in the care of Individuals with Lower Limb Amputation (ILLAs) typically evaluate the patient's physical wellbeing through physical function tests or subjective questionnaires filled out by the patient. These evaluations give a limited understanding of the ILLA's physical wellbeing, which can be evaluated more in-depth via wearable sensor-based Human Activity Recognition (HAR) of physical activities. The key objectives of this thesis were to determine which physical activities could be of interest to HCPs, develop a portable sensor-based system to capture those physical activities, then evaluate how reliably those activities could be captured with wearable sensors.

Methodology: A focus group was conducted with ILLAs and HCPs to identify the relevant outcome measurements for clinical assessment. A novel HAR study was conducted with ILLAs and non-amputated individuals wearing a thigh-bound accelerometer (ActivPAL™, PAL Technologies, Glasgow, UK) to evaluate how reliably these outcome measurements could be captured in free-living conditions.

Results: The key activity monitoring outcomes identified were walking activities on a variety of terrains. Using supervised machine learning, a Support Vector Machine could capture walking activities on flat terrain, walking on hills and walking on stairs. There was further potential to distinguish the activities on walking terrains based on whether they were hard or soft. With unsupervised machine learning, it was possible to distinguish walking on flat or sloping terrain with walking up and down stairs without the need for annotated training data using a novel formula-based algorithm. The ActivPAL proprietary algorithm was also validated for detecting walking and stationary activity of ILLAs in free-living conditions.

Conclusion: The thesis validated an activity monitoring system that could capture a variety of walking activities performed by ILLAs. These findings form the basis of a clinical activity monitoring framework which would allow HCPs to monitor the walking activity of their patients and gain a greater understanding of their rehabilitation progress.

Disclaimer:

This thesis is sponsored by PAL Technologies UK. They provided ActivPAL devices to the author and gave technical advice relating to their operation, however they did not influence the design or outcomes of the research.

Dedication

This thesis is dedicated to Charlie, my best boy.

TABLE OF CONTENTS

CHAPTER

CHAPTER

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Abbreviation	Full Definition	Chapter
ILLA(s)	Individual(s) with Lower Limb Amputation	1
HAR	Human Activity Recognition	1
HCP	Healthcare Professionals	1
TAPES	Trinity Amputation and Prosthetic Evaluation Scale	2
MET	Metabolic Equivalent of Task	2
PASIPD	Physical Activity Scale for Individuals with Physical Disabilities	2
NCPO	National Centre for Prosthetics and Orthotics	3
GPS	Global Positioning System	3
RFID	Radio-frequency Identification	4
ADC	Analogue to Digital Conversion	4
IMU	Inertial Measurement Unit	4
EMG	Electromyography	4
SFS	Sequential-Feature-Selection	4
PCA	Principal Component Analysis	4
LDA	Linear Discriminant Analysis	4
kPCA	Kernel Principal Component Analysis	4
tSNE	t-distributed Stochastic Neighbour Embedding	4
UMAP	Uniform Manifest Approximation and Projection	4
SVM	Support Vector Machine	4
kNN	k-Nearest Neighbour	4
FFT	Fast Fourier Transform	4
CPU	Computer Processing Unit	4
DT	Decision Tree	4

RF	Random Forest	4
AB	AdaBoost	4
BN	Bayesian Networks	4
DAG	Directed Acyclic Graphs	4
NB	Naïve-Bayes	4
DBSCAN	Density-Based Spatial Clustering of Applications with Noise	4
GMM	Gaussian Mixture Models	4
EM	Expectation Maximization	4
FF	Feedforward	4
MLP	Multi-Layer Perceptron	4
CNN	Convolutional Neural Network	4
RNN	Recurrent Neural Network	4
LSTM	Long-Short Term Memory	4
GTA	Ground Truth Annotation	5
VoTT	Visual object Tagging Tool	5
TV	Total Variation	5
AP	Anterior-Posterior	5
SI	Superior-Inferior	5
MFCC	Mel-Frequency Cepstral Coefficients	5
GRF	Ground Reaction Force	5
SMOTE	Synthetic Minority Oversampling Technique	6
mrMR	Minimum Redundancy Maximum Relevancy	6
LOSO	Leave One Subject Out	6
IDE	Integrated Development Environment	6
NMI	Normalized Mutual Information	7
ICC	Intraclass Correlation Coefficients	8

FAQ	Frequently Asked Questions	9
GUI	Graphical User Interface	9
GA	Genetic Algorithm	9
API	Application Programming Interface	9

Chapter One

Amputation, Rehabilitation and Physical Activity

1.1 Introduction to Limb Amputation

Limb amputation is arguably one of the most significant adverse events a person can experience in their lifetime. Alongside loss in mobility and dexterity, limb amputation has associations with a considerable number of detrimental physical and psychological conditions: chronic residual limb pain has been found in 78% of amputees 2 years after amputation (Hanley et al., 2007). They also risk suffering from deep vein thrombosis and stump contractures without proper care and rehabilitative steps (Ghazali, 2018; Yeager et al., 1995). Limb amputation has strong associations with mental disorders including depression (recorded rates as high as 60%) and post-traumatic stress disorder (as high as 56%) (Margoob et al., 2008; Muzaffar et al., 2012; Sahu et al., 2016). Phantom limb pain, the perception of pain in the missing limb, is also a prevalent disorder with amputees, occurring at rates between 45-88% (Ahmed et al., 2017; Wall, Novotny-Joseph, and Macnamara, 1985). Ergo, limb amputation is typically seen as a last resort if needed as a life-saving procedure or to avoid chronic limb pain. (Gardner et al., 2011; LaPlante, 2016).

Table 1.1 Primary causes of Limb Amputation in Scotland, 2011 to 2016, recreated from Davie-Smith, Hebenton, and Scott (2018)

Aetiology of Amputation in Scotland (%)	2011	2012	2013	2014	2015	2016
PAD without diabetes	44.9	41.7	41.3	39.3	40.6	37.1
Diabetes	39.1	42	43.7	46.5	44.7	49.8
Trauma/Burns	1.6	2.7	1.6	2.1	2	1.3
Tumour	1.7	1.4	1.6	2	1.1	1.3
Congenital deformity	0.6	0.4	0.3	0.6	0.7	0.3
Drug Abuse	1.5	1.7	1.6	1.7	2.4	2.1
Other	10.6	10.1	9.9	7.8	8.5	8.1
Total cases	**688**	**702**	**803**	**812**	**704**	**720**

PAD - Peripheral Arterial Disease

Following information from Table 1.1, the aetiology of limb amputation in Scotland can be broken down into two major components: The most frequently occurring cause of amputation are peripheral vascular diseases, wherein a build-up of plaque in the individual's blood vessels causes obstruction or blockage of blood flow. Peripheral Vascular diseases represent the origin of 37.1% of lower limb amputation cases in Scotland as of 2016. This statistic does not include diabetes, which represent a further 49.8%, an increase of 10% of the proportionate share in just five years. Individuals with diabetes are between five and twenty-three times more at risk of requiring limb amputation than someone without diabetes (Jude et al., 2001; NHS Information Centre, 2011), primarily due to high glucose levels causing peripheral neuropathy in the limbs (Beks et al., 1995). Another common cause of amputation are traumatic or burn-related amputations, wherein significant physical trauma has made the limb non-salvageable, these represent a small percentage of amputation incidents in Scotland (1.3%) but around 16% of amputation cases in the USA (Dillingham, Pezzin, and MacKenzie, 2002). Typically, traumatic amputations are associated with armed combat: of 8058 military casualties in Afghanistan and Iraq between 2001 and 2006, 7.4% had to undergo traumatic limb amputation (Stansbury et al., 2008). Due to United States having 10 times the amount of individuals on active duty compared to Scotland and the United Kingdom (Review, n.d.), this likely explains the significant discrepancy in traumatic amputation cases between the two nations. A considerable number of traumatic amputations have also arisen from machinery accidents, motor vehicle collisions and motor vehicle- pedestrian collisions, almost 9000 traumatic accidents in civilian settings occurred between 2000 and 2004 in the United States (Barmparas et al., 2010). Other causes of amputation cases stem from cancer-related diseases and congenital malformations (Dhammi and Kumar, 2014; Dillingham, Pezzin, and MacKenzie, 2002; Montesinos-Magraner et al., 2016).

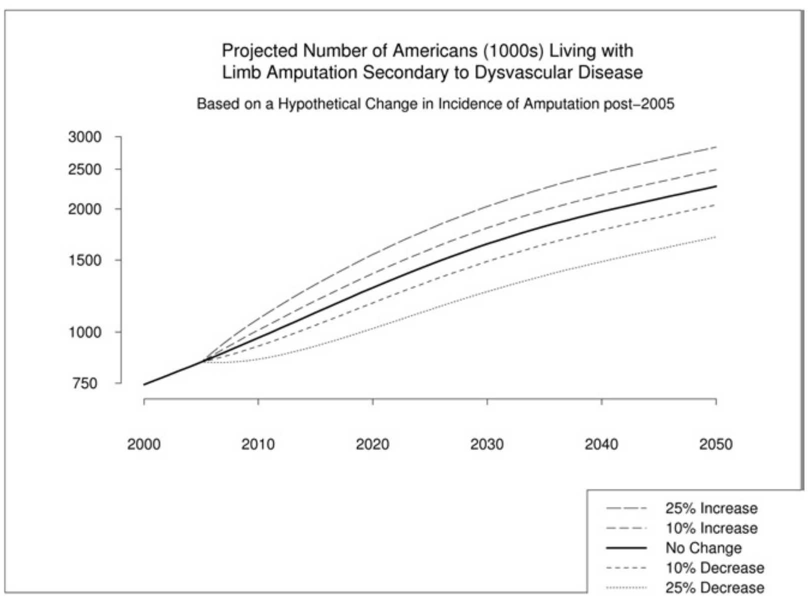


Figure 1.1 Projected number of lower limb amputees by the year 2050. Graph reproduced from (Ziegler-Graham et al., 2008) Copyright © 2019 Ziegler Graham et al. Permission for re-use of figure granted by Elsevier and the Copyright Clearance Center.

As shown by Fig. 1.1 the number of limb amputations in the United States is projected to be 1 in 85 by the year 2050, and of all amputation cases, the majority (65%) of those cases are lower limb amputations (Ziegler-Graham et al., 2008). This statistic appears to have held true as late as 2017 in a study conducted by (Harding et al., 2020). Lower limb amputation can be either unilateral (one leg is amputated) or bilateral (both legs are amputated), with the majority (around 79%) of lower limb amputations being unilateral (Davie-Smith, Hebenton, and Scott, 2018). The amputation procedure can then be further broken down into transtibial amputation (amputation through the tibia) and transfemoral amputation (amputation through the femur), with the ratio between the two types approximately even (Davie-Smith, Hebenton, and Scott, 2018). Other types of amputation procedures exist such as transpelvic amputation and hip/knee/ankle disarticulations. Disarticulations are a rare sub-category of amputation (Davie-Smith, Hebenton, and Scott, 2018), that allow partial preservation of the limb (for example, conserving bone and muscle tissue), as opposed to the complete removal of the affected limb during standard amputations. As a result, the disarticulation operation is considerably less traumatic compared to a standard amputation (where all muscle and bone at the amputation level must be severed) and individuals who experience disarticulation will have less reduction in mobility, especially in ankle and knee disarticulation (Baumgartner, 2011; Saleme et al., 2013).

1.2 Rehabilitation of Lower Limb Amputees

After lower limb amputation, the individual will undergo a rehabilitative process, where the objective is to maximize the patient's functional mobility, independence, and quality of life, such that the adverse impacts of lower limb amputation are mitigated. Esquenazi (2004) gave an overview of the phases of the ideal rehabilitative process an ILLA may undertake. Due to social or financial factors, many ILLAs will only experience a fraction of these rehabilitative processes:

1. **Preoperative**: Prior to surgery, the first stage involves assessing the individual regarding which operative procedure needs to be taken. Preoperative rehabilitative actions

can also be taken, such as strength training led by physiotherapists, and evaluation of potential rehabilitative outcomes led by rehabilitative psychologists (Dekker et al., 2018).

2. **Amputation Surgery**: The patient undergoes appropriate amputation of the leg.

3. **Acute Post-Surgical**: The patient is given treatment to heal their stump wounds, minimize post-operative pain, and will be given emotional support if needed. This is an important stage for managing post-operative pain, as high levels of pain during this phase is often associated with high levels of chronic post-surgical pain (Srivastava, 2017).

4. **Pre-prosthetic**: The individual's stump is prepared for the fitting of a prosthesis. This is also when the multidisciplinary team will aim to restore the individual's *"locus of control"*, meaning that they will encourage the individual to believe in their own ability to retain control over their life. One of the ways of doing so is by giving the individual an immediate post-operative prosthesis which are often used to assess which kinds of mobility assistance devices will be required post-amputation (Ali et al., 2013), however patients in this phase are mostly sedentary (Samuelsen et al., 2017).

5. **Prosthetic Prescription**: The multidisciplinary team will determine which type of prosthesis is best suited to the individual and fit the prosthesis. This decision is generally based on the patient's 'K-level' (Medicare, n.d.) (via (Balk et al., 2018)), which dictates the individual's potential to function with a prosthesis. The K-level, which is ranked from K-0 (does not have the ability to ambulate with a prosthesis) to K-4 (can ambulate with prosthesis and can perform vigorous activities such as running) is calculated based on a variety of factors. Factors include any comorbidities that can impact on the patient's mobility and their own goals for ambulation (Passero, 2014). The team may also make use of mobility predictors, such as the Amputee Mobility

Predictor to inform their decisions (Gailey et al., 2002).

6. **Prosthetic Training**: The patient will undergo training with their prosthesis to familiarize themselves with their prosthesis and to understand their functional capabilities and limits. Prosthetic training at early stages of fitting is crucial for improving physical function and to score well in performance-based metrics (Christiansen et al., 2015). Extended prosthetic use is also correlated with decreased absence of phantom limb pain (Raichle et al., 2008).

7. **Community Training**: After familiarization with the prosthesis, the rehabilitation team will aim to re-integrate the amputee into the amputee's community and can refer to specialist gym classes and other resources to develop the individual's relationship with their prosthesis. Socialization is an important component to the individual's rehabilitation and has been associated with better ambulation potential and quality of life (Hawkins et al., 2016). ILLAs who participate in their community are also found to fall less frequently than ILLAs who do not participate (Hordacre, Barr, and Crotty, 2015).

8. **Vocational Rehabilitation**: In situations where the individual is no longer capable of working at their occupation pre-amputation, rehabilitation teams will work towards providing resources for getting the individual prepared for new occupations, identifying and tackling any physical or motivational barriers that may come from securing the job.

9. **Follow-up**: Lower limb amputation is irreversible, and so throughout the remainder of their life the individual may meet with their rehabilitation team to fix or change their prosthesis, receive updated assessments on their functional status, or to be provided new information, resources and support.

1.2.1 Limitations in Modern Lower Limb Amputation Rehabilitation

The outline of the rehabilitation process described by Esquenazi (2004) describes the ideal rehabilitative process. As will be seen in chapter 3, many ILLAs do not experience these stages. In the pre-operative stage, many of the processes described by Dekker et al. (2018) are not feasible to carry out; while several of the rehabilitative stages describe the user's journey with a new prosthesis, many ILLAs do not receive a prosthesis at all. In 2016, 40.7% of new lower limb amputee patients were fitted with a prosthesis in Scotland as of 2016 (Davie-Smith, Hebenton, and Scott, 2018), 13.4% of the new amputee patients died within 30 days of their amputation, however the largest proportion of amputees were either not fitted for a prosthesis or abandoned their initial prosthetic treatment (45.8%). The breakdown of factors for non-limb fitting (other than abandonment, which accounted for 6.8% of all "not limb-fitted" patients) are not specified in the survey, however some possible reasons can include financial issues, existing comorbidities and socioeconomic factors (e.g not having a partner to assist with prosthetic training) (Pasquina, Carvalho, and Sheehan, 2015). One of the biggest challenges with rehabilitation processes for ILLAs, and the challenge this thesis will aim to tackle, is how to provide effective evaluation of the individual's rehabilitation progress. A handful of papers and reviews in previous research have called for more effective forms of evaluation of rehabilitation process throughout its stages. (Agrawal, 2016; Deathe, Miller, and Speechley, 2002; Hebert et al., 2009; Resnik, Borgia, and Silver, 2017; Salter et al., 2005). This thesis has chosen to focus on how physical activity outcomes are measured in the rehabilitative process. Primarily, this is because physical activity measurements used in interventions are underdeveloped and underutilized compared to physical function assessments, this reasoning is explained in further detail in chapter 2.

1.3 Physical Activity in Lower Limb Amputees

It is important that physical activity assessment and physical function assessment are distinguished as two separate concepts. The World Health Organization defines physical activity as "any bodily movement produced by skeletal muscles that requires energy expenditure" (World Health Organization, n.d.). Physical function, in the context of the rehabilitation of ILLAs, is the "ability to perform both basic and instrumental activities of daily living" (Garber et al., 2010). Physical function assessments will seek to determine whether the ILLA can perform an action (and the level of difficulty associated with performing it), and physical activity assessments ask about how often that action is performed. Another key difference between the two concepts is that a HCP (prosthetists, specialist physiotherapists, clinicians etc.) can observe physical function through short, standardized tests (for example the 2 Minute Walk Test (Brooks et al., 2001)), but cannot physically observe physical activity, which is inherently a long-form evaluation process.

Physical activity for the general population is universally recommended and has overwhelming evidence of health benefits (Reiner et al., 2013; Taylor, 2014; Warburton, 2006). This extends also to ILLAs; physical activity has evidently improved heart and lung functionality and reduced the effects of chronic lower back pain (Bragaru et al., 2011; Shin et al., 2018). From a psychological perspective, physical activity has improved perceptions of the individual's quality of life, self-esteem, and body image (Deans et al., 2012; Wetterhahn, Hanson, and Levy, 2002).

With changes in mobility and lifestyle, ILLAs are unsurprisingly less physically active than individuals without limb loss (Bussmann, Grootscholten, and Stam, 2004; Langford et al., 2019), and there has been some established research into the physical and socio-economical barriers that prevent ILLAs from performing physical activity (Littman et al., 2014; Deans et al., 2012; Bragaru et al., 2013). Therefore, HCPs should monitor and evaluate the physical activity of their clients. Analysing the processes of monitoring activities carried out by ILLAs and how these processes can be improved is the overarching theme of this research work.

1.4 Thesis Overview

1.4.1 Aims and Objectives

The primary aims and objectives of this thesis are:

- Review clinical procedures of physical activity monitoring and evaluation and identify key areas where improvements/progressions must be made
- Conduct an investigation to identify important activity monitoring outcomes for physical rehabilitation of Individuals with Lower Limb Amputation (ILLAs)
- Create a novel Human Activity Recognition (HAR) study to capture physical activities carried out by ILLAs
- Perform machine learning-based analysis of the HAR study and translate the analysis into clinically-relevant information
- Propose a framework for future clinical activity monitoring of ILLAs

1.4.2 Summary of Structure

The first chapter of the thesis, *"Amputation, Rehabilitation and Physical Activity"*, is an introduction to the concepts of lower limb amputation, rehabilitation and how the rehabilitative process of the ILLA ties in with the importance of physical activity.

The second chapter, *"Interventions of Physical Activity for Individuals with Lower Limb Amputation: A Systematic Literature Review"* is a literature review of interventions that have been deployed to try and improve the diversity and frequency of physical activities carried out by ILLAs. In the chapter, a key finding is that interventions to improve physical activity in ILLAs are limited by simplistic activity monitoring outcomes, which leads to a focus on the types of activities that are being assessed by clinicians specialising in the care of ILLAs.

The third chapter, *"Qualitative Analysis of Activity Monitoring Outcomes for Individuals with Lower Limb Amputation through Focus Group and Interview Sessions with Stakeholders"*, is a study of a series of interviews and a focus group that were conducted with ILLAs and Healthcare Professionals (HCPs) specialising in the care of ILLAs. The purpose of the interviews and the focus group was to establish how ILLAs are evaluated for their rehabilitative progress in a clinical environment, establishing perspectives from both the HCPs and the ILLAs separately. The chapter also establishes the key physical activity monitoring outcomes that could be realistically monitored and applied in a clinical setting. Because of the focus on identification of different types of physical activity, the proposed solution is a HAR study which incorporates the activity monitoring outcomes established in this chapter.

As contemporary approaches to HAR are primarily embedded in artificial intelligence and machine learning techniques, the fourth chapter, *"Building Towards a Novel Solution: A Review of Human Activity Recognition"*, is another literature review, this time detailing various practices used in HAR studies and uses the literature review to justify and establish a novel HAR study of ILLAs and non-amputees in free-living conditions using a single wearable accelerometer.

The fifth chapter, *"Collection of Detailed Physical Activity Data from Non-Amputated Individuals and Individuals with Lower Limb Amputation in Free Living Settings"*, constructs the key methodology of the HAR study. It details the participants that were recruited, the key pre-processing techniques required, the features calculated to be fed into the classification models, and outlines the machine learning analysis that is then subsequently conducted in the following chapters.

The sixth chapter, *"Supervised Learning of Human Activity Recognition for Non-Amputated*

Individuals and Individuals with Lower Limb Amputation in Free-Living Conditions", analyses the data collected from the HAR study from a supervised learning perspective, obtaining the most appropriate supervised classifiers to perform HAR on the dataset, experiments with various levels of supervised label detail and investigates the potential for transfer learning by training the classifiers on non-amputees and testing on the ILLAs.

The seventh chapter, *"Unsupervised Cluster Analysis of Walking Activity Data for Non-Amputated Individuals and Individuals with Lower Limb Amputation "*, complements the preceding chapter with an unsupervised analysis of the dataset, aiming to establish identification of physical activities in ILLAs and non-amputees without the use of labelled data to train the classifiers.

The eighth chapter, *"Exploration of ActivPAL hardware and software for further development of activity monitoring capabilities for Individuals with Lower Limb Amputation"*, explores the capabilities of the wearable accelerometer that was used in the HAR study and its proprietary activity monitoring algorithm. The first experiment tests the accuracy of step count of lower limb amputees in free-living conditions using the proprietary algorithm, the second experiment again validates the algorithm but with its capability of detecting events when the participant in the study is stationary. The final experiment looks at the potential applicability of the sensing device's magnetometer for improving performance in machine learning analysis, both from a supervised and unsupervised machine learning perspective.

The concluding chapter, *"Implications of Research in Clinical Context & Retrospective Evaluation"*, coalesces all information gathered from the previous chapters and proposes the outline for a clinical activity monitoring framework, in which HCPs would be able to track the physical activity progress of their ILLA clients. This is followed by a comprehensive summary of the research that had been performed in the thesis, and a discussion as to what the direction of future research should take.

1.5 Accomplishments

The main accomplishments of the thesis are summarized as follows:

- The research constructs the framework for a clinical activity monitoring system in a logical order; instead of beginning with a HAR study and then retroactively try to understand the clinical significance of the findings, the research begins with an understanding of the needs of HCPs and ILLAs, and constructs the HAR study based on those findings.
- The research direction was adapted to cope with the coronavirus epidemic during the majority of 2020 and part of 2021. The adaptation to conduct HAR in free-living conditions also helped contribute to the novelty of the research, as few HAR studies involving the ILLA population have conducted their investigation outside of a controlled laboratory setting.
- The supervised analysis of ILLA physical activity makes use of experimenting with various levels of labelling "resolution" (the detail of the annotated label) and conducts a transfer learning experiment to determine the cross applicability of non-amputee data to train a classifier to recognize ILLA data. Both concepts have been rarely explored in HAR literature, and have never been combined together in a single study.
- The unsupervised analysis of ILLA physical activity is entirely novel. Additionally, a specialised algorithm was theorized and created to help identify physical activities without the need of annotated training data
- The additional experiments conducted in Chapter 8 are, while niche, still novel investigations that have not been conducted before, specifically on the wearable device that was used in the HAR study.

- The research has led to the culmination of two published articles, with another published article to hopefully follow after the completion of this thesis. The research has also been featured at a conference. The following section gives specific details of how the research has contributed to the academic field.

1.6 Academic Output

Publications

- Jamieson, Alexander, Laura Murray and Arjan Buis (2020). *"The use of physical activity outcomes in rehabilitation interventions for lower limb amputees: a systematic review"*. In: Canadian Prosthetics & Orthotics Journal 3.1. DOI: 10.33137/cpoj.v3i1.33931
- Jamieson, Alexander et al. (2021) *"Human Activity Recognition of Individuals with Lower Limb Amputation in Free-Living Conditions: A Pilot Study."* Sensors. 2021; 21(24):8377. DOI: 10.3390/s21248377
- (Pending Review) Jamieson, Alexander et al. (2022) *"Unsupervised Cluster Analysis of Walking Activity Data for Healthy Individuals and Individuals with Lower Limb Amputation"* IEEE Transactions of Biomedical Engineering

—

Conference Contributions

- Jamieson, Alexander, Laura Murray and Arjan Buis (2021) *"Enrichment Of Wearable Sensor Data From Individuals With Lower Limb Amputation In A Free-Living Setting" International Society for Prosthetics and Orthotics, 18th World Congress*

—

Datasets

- Jamieson, Alexander (2021). "Dataset for: Construction of a Clinical Activity Monitoring Framework Based on Free-living Investigations of Individuals with Lower Limb Amputation". DOI: 10.15129/ae451315-5258-4a07-8eb4-204e4d2e357f (University of Strathclyde PURE Open Access Dataset)

1.7 Chapter Conclusion

This first chapter is written to introduce the drive of the thesis, that is, to improve the physical well-being of ILLAs. This began with an overview of lower limb amputation and how it has become increasingly prevalent, showing there is a need for strong rehabilitative services. The rehabilitative process was then linked to the importance of physical activity in ILLAs. The succeeding chapter will explain in a systematic literature review how monitoring techniques of physical activity have been applied in the context of physical and behavioural interventions to improve the physical activity of ILLAs, this chapter is also where limitations to activity monitoring techniques of ILLAs are highlighted, which then creates a justification for the research that is executed for the rest of the thesis. The thesis as a whole is also summarized, along with the publications that have been created from this research.

Chapter Two

Interventions of Physical Activity for Individuals with Lower Limb Amputation: A Systematic Literature Review

2.1 Overview

This chapter is an adaptation of a systematic literature review the author has previously published in the Canadian Prosthetics and Orthotics Journal, (Jamieson, Murray, and Buis, 2020). The journal allows re-use of the manuscript in the author's own works without permission, provided explicit citation to the original journal is given. The original focus of this thesis was to investigate interventions of physical activity and propose how these interventions could be improved. However, as the thesis developed, the theme of the research pivoted towards the measurements of physical activities used in the interventions of themselves, and how they could be applied in an activity monitoring system used in a clinical environment. Thus, this chapter is an abridged version of the author's original published work, restructured to fit the main narrative of the thesis.

Interventions which have focused on improving the physical activity of individuals with lower limb amputation can be broken down into two major categories: prosthetic interventions and behavioural interventions. In a prosthetic intervention, the subject is fitted with a prosthetic component, and their physical activity is typically compared with subjects wearing a variant of that prosthetic component. Marked improvements in physical activity indicate that the prosthetic intervention has helped the patient carry out more physical activity, whether by making them feel more comfortable wearing the prosthesis, reducing the socket pain, or wearing during gait, or any other number of potential physical or psychological factors. A behavioural intervention on the other hand will aim to employ behavioural change techniques such as goal setting, self-monitoring of behaviour and behaviour substitution to the subjects (Michie et al., 2011), which can then be measured in quantifiable activity, such as the number of steps taken per day (Yamada et al., 2012). Other categories of physical activity interventions exist, such as massage interventions, where a therapeutic massage targeted at lowering limb and back pain is applied to an individual and activity levels are measured before and after the massage (Larson, 2015), however the paucity of these interventions makes them unsuitable for the scope of this review.

2.2 Aims

The two primary aims of this review were to assess the quality of prosthetic and behavioural interventions when they are used to modify physical activity behaviour of physical activity performance in ILLAs, while also understanding the ways in which these activities are assessed by the use of physical activity monitoring devices and surveys. The research in this study was important to assess the current state of behavioural interventions and prosthetic interventions in how they modify the physical activity behaviour of ILLAs.

2.3 Summary of the Main Findings

For a full explanation of the methodology and results from this literature review, the reader is encouraged to read the original source material in Jamieson, Murray, and Buis (2020). To briefly summarize the methodology and results: the databases of Scopus, Pubmed, Embase, Medline and Web of Science were searched for articles relating to physical activity, amputees, and interventions. All levels of lower limb amputation were included , so long as the subjects utilised a prosthesis or other walking support devices and were not exclusively wheelchair bound. Only studies that were available in full text and in the English language were considered for inclusion, but no limits were placed on publication years. Articles were assessed quantitatively based on internal validity, external validity, and intervention intensity. In total, sixteen articles were assessed; 5 focusing on behavioural interventions, and 11 on prosthetic interventions. Both approaches had varying methodological quality and mixed efficacy when it came to producing a significant change in physical activity outcomes. After all identified literature were assessed for their internal validity, external validity, and intervention intensity, it was found that behavioural and prosthetic interventions had roughly equal efficacy when it came to generating a significant change in physical activity behaviours. Statistically, the mean scores of internal validity, external validity and intervention intensity were equal between the two groups. The key finding that drove the subsequent direction of the research carried out in this thesis, was that almost all interventions used unreliable (e.g non-standardized questionnaires) or basic activity measurements (e.g step count, activity count) as their outcome measure.

2.4 Behavioural Interventions to moderate Physical Activity in ILLAs

With behavioural interventions, only two studies identified; Christiansen et al. (2018) and Ploeg et al. (2006), had significant positive increases in physical activity behaviour in regards to daily step count, sport participation and the ability to meet pre-defined physical activity requirements. It is also important to consider that the findings of Ploeg et al. (2006) had questionable impact on ILLAs, as they only report their intervention's impact on the general disabled population. Another behavioural intervention, Delehanty and Trachsel (1995) had a single positive increase in "activity" (increased holiday time) while the rest had no significant results. These findings differentiate from reviews which have looked at behavioural intervention studies for people with non-specific disabilities; Castro et al. (2018) and Lai et al. (2017) found significant positive increases in physical activity outcomes in 70% and 83% of identified studies respectively. The meta-analysis used in Ma and Martin Ginis (2018) reported "small to medium sized effects" in the interventions towards physical activity outcomes. A possible explanation for these differing results is the lack of available studies relating specifically to ILLAs: compared to the five articles found in this review, 38, 132 and 24 studies were identified in Castro et al. (2018), Lai et al. (2017), and Ma and Martin Ginis (2018) studies respectively. Another possible explanation is that behavioural interventions may need to tailor the intervention around solving the ILLAs' barriers to physical activity, such as those identified in Littman, Bouldin, and Haselkorn (2017). Despite the lack of evidence and the mixed results, there is some optimism in these findings; by considering that the more modern interventions applied in Christiansen et al. (2018) and Littman et al. (2019) had higher methodological quality than the older interventions, it is possible that future studies will retain a similar high level of methodological quality, which could lead to a more conclusive idea of how effective behavioural interventions are on the physical activity of ILLAs in the future.

2.5 Prosthetic Interventions to moderate Physical Activity in ILLAs

Prosthetic interventions also had mixed effects on the physical activity of ILLAs, with five out of twelve studies reporting significant effects. This finding is echoed by Samuelsson et al. (2012) and Pepin, Akers, and Galen (2018) who both reviewed the effects of prosthetic components on physical activity. In Samuelsson et al. (2012) and Pepin, Akers, and Galen (2018), five out of eight studies and five out of fourteen studies had significant impact on physical activity outcomes respectively. The findings of the review are comparable to Samuelsson et al. (2012) as they used the same reviewing criteria (internal and external validity) and reviewed some of the same articles. The external validity was found to be scored identically in each of the shared articles, however there were some minor disagreements with internal validity criteria and scoring. For example, in the assessment of Coleman et al. (2004) they scored 0 for reporting psychometric properties of the measuring instrument, while this review scored a 1. These discrepancies can be explained by the differing objectives that the review by Samuelsson et al. (2012) had. In Coleman et al. (2004) the psychometric properties of the physical activity measuring instrument were reported, but not the questionnaires. As these questionnaires report on the impact of quality of life and participation in the individual's community, which were key topics in the review by Samuelsson et al. (2012), this likely explains why Coleman et al. (2004) scored a 0 in their review for that particular element. The maximum discrepancy in internal validity scoring was ± 1, so overall both reviews had a similar assessment of the shared articles.

Only one prosthetic intervention to moderate physical activity had been developed in the time between the review by Pepin, Akers, and Galen (2018) and this review. Considering this finding, it appears that there have been minimal adoption or development of prosthetic interventions to moderate physical activity outcomes.[1] At best, they appear to have mixed

[1] Author's note: This observation was made in early 2020, prior to the coronavirus outbreak

efficacy, and even within the same intervention type, results are inconsistent. For instance, all identified prosthetic knee interventions compared a microprocessor knee to a mechanical knee, and multiple outcomes were found; two papers reported no significant results in activity outcomes (Hafner et al., 2007; Klute et al., 2006), one reported significant improvement in favour of wearing the microprocessor knee (Kaufman et al., 2008), and one reported significant improvements in favour of wearing the mechanical knee (Theeven et al., 2012). The review therefore concludes that prosthetic interventions are, in their current state, an unreliable method of improving physical activity outcomes. Some promising developments in prosthetic technology could be incorporated into the design of future prosthetic interventions. For example, powered knees are a recently developed type of prosthetic knee that, compared to the more traditional microprocessor and mechanical knees, provide greater output in energy assistance and can help perform more demanding walking movements like climbing stairs (Pasquina et al., 2015). These inventions may be critical to obtaining definitive improvements in physical activity behaviour in ILLAs.

2.6 Lack of Identifiable Studies that use Physical Activity Outcomes

A significant finding obtained from this systematic review is that physical activity-based outcomes for ILLAs are rarely utilised in interventions, which may stem from their lack of prevalence in clinical practice. In a survey of which outcome measures were used by prosthetists and orthotists to assess their patients, a single questionnaire to assess measurement of physical activity (the Trinity Amputation Prosthetic Evaluation Scale (TAPES)) appeared on a list of the most frequently used outcome measures (Young, Rowley, and Lalor, 2018). The driving factor of these findings may come from the fact that two of the major ILLA rehabilitation organizations – the British Association Of Chartered Physiotherapists In Amputee Rehabilitation, and the British Association of Prosthetists and Orthotists – do not list any physical activity-based outcomes aside from TAPES in their outcome measurement guides (Scopes et al., 2015; Young et al., 2015). The majority of outcome measures used are functional/mobility tests (e.g Timed Up and Go test (Podsiadlo and Richardson, 1991)) and physical function scales (e.g Houghton Scale (Devlin et al., 2004)). These types of measurements have their limitations; while physical function has been shown to have a positive correlation with physical activity (Manini and Pahor, 2008; Metti et al., 2018; Yorston, Kolt, and Rosenkranz, 2012), some criteria used in physical function scales relies on either subjective assessment from the assessor, or self-assessment from the patient (e.g in the Houghton scale, patients are asked how many hours per day they wear their prosthesis), which can lead to recall bias and unreliable results. While mobility tests use quantitative outcomes (e.g the Timed Up and Go test is assessed in terms of the amount of time taken to perform the test), they are only a "snapshot" of the patient's well-being. The ILLA's performance in the test could fluctuate depending on external and environmental factors, for example how tired the patient is, or the type of footwear they are wearing can have an impact (Sprint, Cook, and Weeks, 2015). Activity measurement outcomes on the other hand benefit from being a much more dynamic assessment of the ILLA's physical well-being. These established outcome measurements should not be replaced with physical activity outcomes altogether, but rather they should supplement the existing methodology.

2.7 Outcome measures of physical activity used in behavioural and prosthetic interventions

In the behavioural approach, two interventions used objective activity monitoring measurements (Christiansen et al., 2018; Littman et al., 2019), two interventions used subjective questionnaires (Kosma, Cardinal, and McCubbin, 2005; Ploeg et al., 2006), and two interventions used non-standardized questionnaires (Delehanty and Trachsel, 1995; Ploeg et al., 2006). By contrast, all prosthetic interventions used objective activity measurements.

Subjective measurements of physical activity are primarily composed of questionnaires and diaries, where the ILLA is fully or partially responsible for judging their own physical activity. These techniques are popular in clinical practice as they are often inexpensive to adapt, and in the case of self-reporting questionnaires, there are an abundance of standardized questionnaires that healthcare professionals can utilise. Most ILLA-specific questionnaires will specifically focus on physical function and the user's comfort with their prosthesis, for example, the Prosthetic Evaluation Questionnaire and the Locomotor Capabilities Index (Boone and Coleman, 2006; Franchignoni et al., 2004; Legro et al., 1998). Prior to the turn of the century, questionnaires to evaluate physical activity of ILLAs were non-standardized and uninformative; Delehanty and Trachsel (1995) created a "Rehabilitation Status Questionnaire", which prior to the investigation had only been validated with a pilot study. Their outcome measures - which included "Church", "Shopping" and "Banking" – are outdated by modern standards. In Ploeg et al. (2006), sport score and sport participation were assessed by a custom questionnaire which took into account the number of hours spent on the sport and the designated intensity of the sport in Metabolic Equivalent of Tasks (METs) from a physical activity compendium (Ainsworth, 2008; Bouchard et al., 1983). The authors did not provide further details of which sports were carried out and for how long, so it was impossible to identify which activities the ILLA population were participating in. These non-standardised forms of evaluation make it difficult to compare results across different studies and should be avoided in future investigations. Additionally, evaluating ILLA physical activity with METs can be problematic, as METs must be readjusted to account for the fact that an ILLA will require more energy to carry out activities at the same intensity as a non-amputee (Littman et al., 2014).

There are a handful of standardized questionnaires that are specific to measuring the activity of ILLAs, for example certain subsections of the TAPES questionnaire (Gallagher and Maclachlan, 2004; Washburn et al., 2002). Ploeg et al. (2006) and Kosma, Cardinal, and McCubbin (2005) made use of the Physical Activity Scale for Individuals with Physical Disabilities (PASIPD) questionnaire to evaluate their programs (Washburn et al., 2002). PASIPD is a widely used and validated questionnaire (Ploeg et al., 2007). The questionnaire assesses physical activity by combining the number of hours spent performing a particular activity with the activity's MET score. Despite the questionnaire's popularity, the PASIPD has been found to show poor correlation with objective physical activity measurements (Van Den Berg-Emons et al., 2011), and so by itself could be considered an unreliable tool for fitness assessment. Supplementing these questionnaires with objective measurements could be a way of providing a complementary or contrasting viewpoint. For instance, scoring high on a questionnaire while performing poorly on objective measurements could indicate the patient is overestimating their fitness levels.

While not used in any of the identified articles, physical activity diaries can occasionally be used as a tool to record and promote physical activity in other vulnerable populations, such as those with diabetes (Bravata et al., 2007; Furber et al., 2008; Quirk, Glazebrook, and Blake, 2018). A flaw that is common to both subjective questionnaires and activity diaries is the reliance on the individual for measurement. As will be discussed further in Chapter 3.2.2.3, some individuals are prone to exaggeration about their activity levels, even with non-malicious intent, especially if they perceive achieving high levels of activity with receiving praise or access to better clinical resources like a high-tech prosthesis.

The behavioural intervention studies from Christiansen et al. (2018) and Littman et al. (2019) and all included prosthetic intervention studies all used objective activity monitoring, wherein monitoring of physical activity is performed by an unbiased source - a sensor (or sensory array). By far the most common approach for objective physical activity monitoring was to utilise the Step Activity Monitor and then analyse the intervention by changes in some measurement of step activity. Other devices such as the ActivPAL (PAL Technologies, Glasgow, UK) and ActiGraph (ActiGraph, FL, USA) were also used but only to measure step count or vaguely defined ‘activity bouts’. While objective activity monitoring is much more reliable than self-report questionnaires in terms of accuracy (Stepien et al., 2007), monitoring devices are reliant on using step count as their primary or only unit of measure. High step counts have strong associations with positive health outcomes such as a decrease in the risk of cardiometabolic adverse events (Schmidt et al., 2009), and so while they do provide valuable indications of physical activity, they only give surface-level insight – for instance, an ILLA who performs stationary exercises and stretches will appear to be inactive when monitored by an ordinary pedometer. Kaufman et al. (2008) was the only study to measure energy expenditure via the Doubly-Labelled Water Effect. While its high precision makes the this method the gold standard for measuring energy expenditure (Berman et al., 2019), a limitation of this method is its complexity – the method requires ingesting an isotope which is then expunged through urination and analysed using mass spectroscopy. Analysis must be carried out by a specialist, making it impractical to use for large sample sizes. Another problematic issue is that there is no standardisation of energy readings applicable to ILLAs like METs are to non-amputees. Using standard METs to assess non-amputees gives an unequal comparison due to lower energy expenditures(Littman et al., 2014), and bodies such as the American College of Sports Medicine have yet to establish an equivalent system for ILLAs. Likewise, while there are government funded documents such the UK Chief Medical Officers’ Physical Activity Guidelines to help set standards of physical activity for the general population (Davies et al., 2019), there is no equivalent document for ILLAs.

2.8 Future of Interventions for ILLAs

2.8.1 Behavioural Interventions

In a currently ongoing behavioural intervention for ILLAs, titled “Physical activity behaviour change for older veterans after dysvascular amputation”, veterans with a dysvascular amputation will undergo a 3 month long behavioural intervention (Miller et al., 2017)[2]. The intervention will make use of telerehabilitation technology to allow participants to communicate with their physiotherapists via live video communication. In these video sessions, the physiotherapist will discuss barriers to physical activity, educate the participant on activity-related topics such as fall preventions and dieting and record adverse effects. A FitBit® (Fitbit, CA, USA) device will monitor their physical activity and feed the information back to a mobile tablet application, which will help the participant monitor their activity and set out weekly activity goals. While the primary outcome of this trial is to measure its feasibility, physical activity outcomes in terms of physical activity count (measured by an ActiGraph device) will be assessed as part of calculating the effect size of the intervention. When considering this study along with the two most recent studies identified for this review (Christiansen et al., 2018; Littman et al., 2019), it appears that communication technology is playing an increasingly important role in these types of intervention. By providing a medium in which the individual and their healthcare provider can frequently communicate, whether through telephone, video conferencing or e-mail, the individual is given many more opportunities to identify and overcome barriers to their physical activity, and the healthcare provider is given more oversight into their patient's physical activity performance. The use of modern technology to provide behavioural interventions to increase physical activity in other low activity populations has seen some success; for instance a novel web-based behavioural change intervention called e-Motion improved bouts of MVPA in depressed populations (Lambert et al., 2018). A mobile behavioural change application designed to self-monitor weight, caloric intake and physical activity helped improve the step count in diabetic individuals (Fukuoka et al., 2015). In the future, technologically advanced behavioural interventions should start to include or specialise in ILLAs.

Above all, more behavioural interventions are needed to properly assess their effectiveness on ILLAs. This sentiment is mirrored by Jayakaran, Perry, and Hale (2019) which compared

[2] This study is still on-going as of January 2022

physical activity levels of ILLAs with their quality of life: they state that:

> *"Person-centred behavioural interventions to increase physical activity levels are needed to decrease the risk for developing long-term co-morbidities and to lessen the effects of co-morbidities already present in this population".*

Evidently, more studies such as the ones identified in this review should be developed in the future, with higher quality approaches to their methodology.

2.8.2 Prosthetic Interventions

As stated in section 2.5, prosthetic interventions to moderate the physical activity of ILLAs have mixed efficacy, and by comparing the outcomes of this review to previous reviews that have covered this subject matter, it appears this type of intervention has stagnated; meaning that their efficacy does not appear to have improved over time. In part, these results may be due to their largely homogenized methodology; almost all studies used crossover design, and the activity measurements carried out did not expand beyond step count, with a few exceptions. Additionally, most studies had a short intervention length, causing them to score poorly in the intervention intensity. Improvements in study outcomes could be obtained by a change-up in the study design, increasing the length of the intervention, or by changing the physical activity outcomes used.

Some promising developments in prosthetic technology could be incorporated into the design of future activity interventions. For example, powered knees are a recently developed type of prosthetic knee that, compared to the more traditional microprocessor and mechanical knees, provide greater output in energy assistance and can help perform more demanding walking movements like climbing stairs (Pasquina et al., 2015). Another more cost-efficient way to potentially improve the efficacy of prosthetic interventions is to introduce behavioural or cognitive elements, which have previously been found to increase physical activity among older adults versus using a singular approach (Morelhão, Oliveira, and Franco, 2016).

The important characteristic that is needed in all future interventions of physical activity, whether behavioural or prosthetic based, is the integration of objective-based physical activity monitoring outcomes. Despite some of the limitations that have been found with objective activity monitoring techniques, they are still the preferable method of ILLA activity evaluation due to the non-biased nature of the technique. ILLAs will always have unique perceptions of their physical activity capabilities; for example an ILLA who walks 500 steps per day could consider themselves to be very active, while another ILLA who walks 2000 steps per day may consider themselves inactive. The introduction of objective monitoring mitigates the personal biases of the amputee as well as any clinician observing them, and provides clearer scope as to how physical activity capabilities in the ILLA can be improved.

2.9 Chapter Conclusion

After conducting a systematic review on Scopus, Pubmed, Embase, Medline and Web of Science, 16 studies were identified which assessed the physical activity of ILLAs after the application of a prosthetic or behavioural intervention. Ultimately, the lack of available studies makes it difficult to comment on the overall efficacy of behavioural interventions on ILLAs, but the increase of quality of the methodology in the most recent studies identified give an optimistic indication that future interventions will have similar levels of methodological quality. There are a substantial number of prosthetic interventions with good methodological quality, however the efficacy of these prosthetic interventions has stagnated, and may require implementing more technologically advanced prosthetic components to obtain a significant change in activity.

Following on from this systematic review, it became clear that in both types of interventions, the types of activity measurements used by the researchers were unreliable, outdated, or overly simplistic. The findings of the systematic review inspired the next stage in the thesis, which was to work with ILLAs and HCPs specialising in the care of ILLAs to find out what kinds of activities are carried out by ILLAs, and which of these activities are important in the context of clinical evaluation. This research was conducted with the implications that future interventions, regardless of whether they are behavioural or prosthetic-based, should incorporate these more sophisticated forms of activity measurement to give a more in-depth assessment of physical activity. These measurements are used to form the basis of a clinical activity monitoring system for analysing the activity of ILLAs.

Chapter Three

Qualitative Analysis of Activity Monitoring Outcomes for Individuals with Lower Limb Amputation through Focus Group and Interview Sessions with Stakeholders

3.1 Purpose and Procedures

This chapter covers a series of interviews and focus groups (referred to collectively as interviews) with two key stakeholders: ILLAs and HCPs specializing in the care of ILLAs. The structure of the chapter is based on guidelines set out by Krueger (2014)'s book: *"Focus Groups: A Practical Guide for Applied Research"*. In an interview setting, the primary aim is to gather a wide variety of insights and experiences to make informed decisions, however it is not intended to make a statistical inference on the preferences of the stakeholders' general population (Asbury, 1995). As outlined in Chapter 2, there was a need to establish detailed outcome measures for the assessment of ILLA physical activity beyond the measurements currently used in literature, and so these interviews were designed to identify these outcome measures and additionally explore how an activity monitoring system for ILLAs could be applied in a clinical environment.

Explicitly, the reason why qualitative interview-based research was carried out in favour of a quantitative approach (for example, a survey or questionnaire), or researching second-hand sources was twofold: primarily, it was vital that this research should receive specific targeted feedback from the primary stakeholders. Qualitative feedback allows for more nuance in dialogue that a quantitative method cannot provide (Al-Busaidi, 2008). By gathering information first-hand instead of second-hand, it ensures that the views held by the stakeholders are up to date. While some studies focus on the barriers to physical activity for ILLAs (Miller et al., 2019), these studies are themed around improving general physical fitness, and not an analysis of specific activities. Improving fitness from a general perspective is inherently good, but targeting the improvement of specific activities can lead to more constructive actions taken by the HCP; for example to traverse stairs, the HCP could recommend a specific set of stretches that could help the ILLA traverse stairs in a safe and non-strenuous manner. Secondly, it was anticipated that, should the research have used quantitative methods (for example, sending out mass-surveys) there would likely have been a significant lack in responses, partly due to the general obscurity of the research topic, as well as the inability to provide an incentive for participation.

The interviews for this research were conducted over the span of 5 months between October 2019 and February 2020. The research was approved by the University of Strathclyde's Biomedical Engineering Departmental Ethics Committee (Ref: Paper DEC/BioMed/2019/268). Notable inclusion criteria were that ILLA participants must have had their amputation procedure performed at least 1 year prior to the interview, and that HCPs must have had at least 2 years of experience working with ILLAs within a professional setting. These criteria ensured that ILLA participants would have had ample time to reflect on their experience as an amputee, and that HCPs had an informed understanding of how their patients behave.

ILLAs and HCPs were recruited through separate approaches, with ILLAs being primarily recruited through flyers and posters, and HCPs being contacted through networks via a third party. Due to overall low recruitment numbers, the methodology adopted convenience sampling and all interested participants were selected for participation.

In total, 8 HCPs and 6 ILLAs were interviewed. The original intention of this study was to carry out all interviews in two distinct sets of focus groups: one for the HCPs and one for ILLAs. However due to a lack of availability and conflicting schedules, most of the participants were interviewed separately via a phone call interview in accordance with their schedules. All 8 HCPs and 2 of the ILLAs were interviewed separately, and a single focus group was carried out with 4 of the ILLAs. Undoubtedly, changing from a focus group format to an interview format will have made some impact. The additional level of inter-participant interaction inherent to a focus group discussion can provide information that would not be found in their individual interactions (Lederman, 1990). A group dynamic allows for the presence of complementary and contradictory views, thereby highlighting the individuality of each person that can be difficult to capture in a dynamic fashion when conducting interviews separately (Lewis, 1992). To give a hypothetical scenario, if one HCP took a particular treatment route to fit an ILLA with their prosthesis, while several other HCPs took a completely different route, it would not be practical on a large-scale implementation (with more of these scenarios) to contact the first HCP again to try and understand why their treatment route was so different. Whereas in a focus group, this contrast would be immediately apparent and could be discussed further, achieving greater depth of knowledge.

Interviewing participants individually, while perhaps not as productive as a focus group session, allows one to avoid common 'pitfalls' associated with the focus group technique. Social loafing is a common phenomenon in focus groups wherein some subjects may fail to answer questions or to simply agree with other participants without providing additional perspective (Asbury, 1995; Piezon and Ferree, 2008). Likewise, interview-shy subjects can sometimes be present in focus groups, who may feel uncomfortable sharing personal information in a group setting (Grbich, 1999). Being the sole subject in an interview can help with these issues; they will be providing unique and unbiased perspectives of their experiences which is uninfluenced by their cohorts' opinions. The inherently more private setting of a one-on-one interview will help participants share their experiences more freely, which benefited the interview setting as the experience of amputation and rehabilitation can often be a sensitive topic (Senra et al., 2012). Likewise in the case of the HCP demographic, some may be more unwilling to divulge patient anecdotes or treatment procedures to their cohorts than they would in a private setting. A further key element to address was that all individual interviews were conducted over the telephone, which comes with its own unique advantages and drawbacks. An obvious disadvantage is being unable to read facial expressions and body language, which can help the interviewer understand the subtext of the participant's words more easily (Aquilino, 1994). On the other hand, by having an additional anonymous barrier between the interviewer and the participant, the participant may feel more comfortable with sharing information with a person who cannot recognize them in a social setting (Hopper, 1992), and some research suggests that the potential loss of data compared to a face-to-face meeting is most often a negligible factor (Novick, 2008). Research has shown that focus groups and one-on-one interviews can both be considered viable depending on the research topic (Heary and Hennessy, 2012). It is beyond the the scope of this chapter to analyse how these differing interview formats may have had a psychological influence on the participants, at least at a level of detail relating to their given answers. Nonetheless it is still worth acknowledging these differences as a potential confounding factor on the findings of the study.

Interviews followed a semi-structured format lasting a length of approximately 45-60 minutes. A pre-determined list of questions was conceived prior to the interview but allowed for branching and unscripted follow-up questions depending on the answers given by the participants. The format of these questions is contained in Appendix A. All participants were further encouraged to ask follow-up questions when they felt it was appropriate, and in the case of the Focus Group, were encouraged to also interact with each other by comparing and sharing their experiences. Participants were informed of the general theme and structure of the interview through a Participant Information Sheet ahead of time. All audio in the interviews were captured with a digital voice recorder, and all participants explicitly gave verbal consent to allow their audio to be used prior to the beginning of their interviews. The audio of the sessions was later transcribed to a word processing document and analysed using qualitative analysis software NVivo (QSR International, Melbourne, Australia) to identify recurring themes in the responses given by the participants.

The findings are thematically split into 3 areas. The first topic tackles the characteristics of the interviewed parties. The second topic is centred around rehabilitation: it looks at how the HCPs structure their rehabilitation, and how the ILLAs experienced their own rehabilitation. The third topic centres around an activity monitoring system and the outcome measures that should be used in this system. The identities for both stakeholders are fully anonymized, and so in this chapter, specific participants will be referred to using a simple codename format: HCPs are all given a "HCP" prefix followed by a number (HCP1, HCP2 ... HCP8). ILLAs use the same convention but use an "ILLA" prefix instead (ILLA1, ILLA2 ... ILLA6).

3.2 Findings and Interpretations

3.2.1 Participant demographics

In total, six ILLAs were interviewed. ILLA1 and ILLA2 were both interviewed individually over the telephone, and ILLA3-ILLA6 were all simultaneously interviewed in a focus group session. 5 of the ILLAs had unilateral amputation, and 1 (ILLA3) was a bilateral amputee. 5 amputations were at the transtibial level, and 1 (ILLA1) was at the transfemoral level.

There were inarguably some location bias, as all 6 participants were recruited from the greater Glasgow area, with 5 being recruited from the National Centre for Prosthetics and Orthotics (NCPO).

All ILLAs had been discharged from their initial rehabilitation and utilized a prosthesis as their primary means of ambulation. Their initial prosthesis fittings ranged between 1965 and 2017, and the demographic could be roughly split into 3 experienced ILLAs (ILLAs that had worn a prosthesis for over 30 years) and 3 inexperienced ILLAs (worn a prosthesis for less than five years). Four out of the six ILLAs had amputation in response to severe leg trauma, while the other two were from vascular related diseases. Further details on the condition surrounding the amputation were not pursued with the concern of making participants uncomfortable about disclosing sensitive or traumatic memories, especially in the focus group setting. Four of the ILLAs had retired, one was a mature student (ILLA3), and the sixth amputee (ILLA4) was still in active employment as an engineer. Of the retired ILLAs, one amputee (ILLA5) was an army veteran, one amputee was a tradesman (ILLA6) who was forced to change their career to acting after the amputation, one was a nurse (ILLA2), and the remaining amputee (ILLA1) did not specify their previous profession.

Eight HCPs were interviewed over the telephone. Two of the HCPs (HCP1 & HCP7) were physiotherapists specialising in ILLAs, while the other six were a mixture of prosthetists and prosthetist/orthotists. Ideally, there would have been more of the former group, as they would be more knowledgeable about the outcome measures that are used to assess the rehabilitation progress of their patients. Nonetheless, the prosthetists were still informed of what their physiotherapist colleagues were doing due to amputee rehabilitation being a multi-disciplinary effort. The HCPs' years of experience ranged from 2.5 years up to 11 years in their current position. When asked about the "typical" kind of ILLA they worked with, most of the HCPs worked with archetypal ILLAs, meaning patients were typically middle-aged (50 years or older) and required unilateral amputations as a result of vascular diseases (Ahmad et al., 2014). HCP6 was the exception, who primarily worked with younger bilateral amputees (specifying a mean age of around 40) who lost their limbs in military-related traumatic incidents. The interviewed HCPs were located with more diversity than the interviewed ILLAs, though still primarily located in Scotland: 3 HCPs (HCP3, HCP4, HCP7) were based in the Greater Glasgow area, while other HCPs were based in Dundee, Fife, and Edinburgh (HCP1, HCP2, HCP8). 2 HCPs (HCP5 and HCP6) were based in separate clinics in the Netherlands.

3.2.2 Rehabilitation of Lower Limb Amputees

3.2.2.1 Rehabilitative Process Structure

All HCPs used face-to-face interactions in their rehabilitation sessions. HCP5 & 6 both utilised phone calls only in situations where they had no contact with their patient in a long time, or if a face-to-face meeting would not be possible. There was a clear preference among all interviewed HCPs that any important information related to their patient's rehabilitation progress should be discussed face-to-face. The frequency of the rehabilitation sessions led by the HCPs trended from high frequency (multiple times per week) in the initial stages of rehabilitation (stages 1-6 of the rehabilitative process as described in Section 3.2.2.1), to low frequency (several times per year) in the later stages of rehabilitation (stages 7-9 of the rehabilitative process). All HCPs claimed to keep in contact with their clients post-discharge for an indefinite period; HCP8 explained their "open-door policy", stating that rehabilitation was a lifelong process, and they were always welcome to return to a clinic if they are having issues with their prosthesis.

In stark contrast, none of the ILLAs interviewed had an active, on-going relationship with a HCP, which would have given some interesting insight into how their relationship with the HCP developed long-term. When asked why they had lost touch with their HCP, a common theme in the responses was that any contact would only be made in response to an emergency, for example if there was an issue with their prosthesis. This finding from the ILLA interviews align with how the HCPs interacted with their clients post-discharge; all ILLAs were involved with a HCP up to the point of being discharged from a rehabilitation centre or hospital with a prosthesis, after which contact became minimal. The period in which the ILLAs underwent rehabilitation ranged from several weeks to over 2 years, and there was a variety of experiences – ILLA6 was only trained and monitored until they reached K1 levels of amputation and could "barely walk", while the veteran amputee with traumatic amputation (ILLA5) had to undergo intensive rehabilitation until they were at K4 levels of ambulation and were fit to serve in the armed services again. All interactions with their HCPs were held in-person, which aligns with the methodology employed by the interviewed HCPs. Post-discharge, few ILLAs sought further rehabilitation by taking up specialist exercise classes at local community/charity centres, while others just tried to resume their daily lives to the best of their ability.

When discussing the ILLAs' overall thoughts and feelings on their rehabilitative progress, the overall language and tone used was framed in a positive manner. There was a clear lack of specificity when recalling their rehabilitation experiences, which was natural given that their primary rehabilitation was often many years or even decades ago. When asked about what parts of the rehabilitative process they felt were good, most framed the entire experience as useful. In fact, all members of the focus group members described their rehabilitative process as being "all essential", though repeated use of the word "essential" may have resulted from social loafing (Pandeirot and Aseng, 2017). One of the interviewed amputees, ILLA1, also independently brought up the word "essential" to describe their rehabilitation. While the tone conveyed by the ILLAs was mostly neutral, the veteran amputee (ILLA5) was notably more enthusiastic about the experience than their other cohorts. Their rehabilitation was much more intensive due to still being a member of the armed forces at the time. They further highlighted that the rehabilitation sessions minimized their leg pain. When asked about negative experiences, ILLA2 was the only participant to describe one, citing hospital ward-based physiotherapy (therapy with a group of patients) as being unhelpful due to the large number of participants. Overall, there was a clear theme of respect towards their rehabilitation programs, which appeared to stem from the ILLA's trust in their HCPs.

Given that the rehabilitation of their clients is their profession, the language used by the HCP when describing the rehabilitative process was unsurprisingly overwhelmingly positive. It was made clear that the rehabilitative process was a multidisciplinary process, and the HCPs all described their unique roles in the process in great detail. It was made clear that the prosthetist HCPs were heavily involved in the earliest stages of the rehabilitative cycle, when the patient is first receiving their prosthesis and afterwards their involvement is reduced, with subsequent interactions mostly dependent on whether there were some prosthetic issues. The physiotherapists generally became involved in the rehabilitation process once the patients began gait training with their new prosthesis. All HCPs seemed satisfied with how often they held sessions with their patients and increasing the frequency of post-discharge interactions outside of emergency situations would not be feasible due to the number of clients they had.

3.2.2.2 Outcome Measurements in Rehabilitation

Given the lack of specificity regarding their rehabilitative experiences, the ILLAs had expectedly forgotten or were non-descriptive regarding how they were assessed during their early stages of rehabilitation. It should be noted however, in the case of the long-time ILLAs, their rehabilitative processes likely preceded widespread implementation of outcome measures; this is indicated by the tested outcome measures listed in Resnik and Borgia (2011), where most outcome measurement tests were developed in the 1990s and 2000s, with no guarantee of these outcome measures being implemented in Scottish healthcare immediately following publication of the sources. No ILLAs were assessed via any kind of wearable sensors (e.g a fitness tracker). The 2 Minute Walk Test, Timed Up-and-Go test and questionnaires with unspecified titles were brought up in discussion. These remarks align with the research carried out in Chapter 2.7, where it was found that functional assessments and subjective questionnaires were far more common for use in outcome measures. ILLA5, the veteran amputee, was required to undergo additional fitness tests to be allowed back into the armed forces, however these assessments were a one-time event with only a pass/fail threshold to describe the outcome.

Likewise, assessments used by the HCPs was comprised of either functional/fitness tests, or subjective questioning. HCP8 explained that asking how a patient "was getting on" was often the extent of depth of their evaluation. HCP4 described their clinic providing microprocessor knees to some patients, which contained embedded step counting. However, the step count was only assessed for validity by the physiotherapists, meaning that they were primarily concerned with the accuracy of the step count measurement, rather than the quantities of the measurement itself. However, the physiotherapists were mainly concerned with he accuracy of the step count, and did not use the measurements to evaluate their client's activity. Outcome measures played a key role in some clinical decisions: HCP8 described how they could use functional assessments to compare how a microprocessor-based prosthetic knee operated in comparison to a hydraulic knee by giving them both knees for six weeks and comparing the outcome measures at the end of those weeks. By extension, no HCPs had used wearable sensors as an assessment. When asked why, the reasoning fell into one of three categories: prosthetists would argue that due to their involvement only being in the preliminary stages of rehabilitation, wearable sensor monitoring was not relevant to their profession. Others stated that due to a lack of financial resources, the logistics of implementation for wearable sensors were too demanding. The remaining groups simply remarked that it was not something they had investigated. While the opinion on wearable sensors was optimistic but cautionary, there was some criticism on the usage of wearable sensors. HCP8 cited the consistency of the device's placement on the body could be an issue regarding accuracy and precision of measurements. Another issue discussed was how the activity monitoring device could be manipulated by the patient (for example by shaking it), thus requiring robust detection algorithms.

3.2.2.3 Setting goals and providing feedback

With both parties, there were no mentions of providing formal exercise targets as a baseline for improving the patient's physical activity. Primarily, the goals set out by HCPs were deliberately designed to be simple, often binary-based outcomes (was goal met or not) that were mutually agreed upon by the HCP and their client. HCP6 for example explained their goal setting was termed as "aspirations", where the client was responsible for setting their own goals, and the HCP would document whether they had met their aspirations over time. HCP6's clinic was responsible for the recuperation of younger traumatic ILLAs based in the military, and so setting these goals would be vital if they intended to continue to serve. In contrast, a handful of HCPs did not utilise goal setting of any kind. HCP2 outright stated that it was impractical to try and set goals for their clients due to overseeing around 300 patients at a time. They further stated that in the past, they had tried to track rehabilitation progress by sending out questionnaires, but the idea failed due to low response rates. When feedback was provided by the HCP, the format of delivery was always primarily verbal, with occasional use of videos or diagrams to assist or enhance the quality of the feedback. Feedback was carried out with lay terms that the patient could understand, but not too simple to the point of being patronizing. Medical jargon was only used when the patient was an experienced amputee, had scientific knowledge or directly requested a more detailed explanation.

As rehabilitation progress was often subjectively judged by the ILLA themselves, the HCPs were asked whether they felt that their clients would overestimate their rehabilitation progress. All HCPs agreed to an extent that exaggeration in self-report assessments was an issue, though some found it less problematic than others; for instance, while HCP6 stated that:

> *"(Exaggerating patients)... is always the biggest problem and challenge (in my line of work)"*

HCP8 on the other hand remarked:

> *"... most of the time any lies would just be (the patient) being economical with the truth."*

The motivations behind self-exaggeration of progress were also discussed, for instance HCP1 stated their patients may lie if they perceive the possibility of having to change prosthesis or to lose certain kinds of benefits. They also stated that motivation and pride can be underlying factors. HCP7 also mentioned the patient might not want to let their HCP down and so may exaggerate to seek approval subconsciously. In retrospect, this could be considered a leading question, as the framing of the question implied that the interviewer was looking for specific examples of clients exaggerating or lying. A more balanced question would be framed as "do you feel you can rely on your patients to tell the truth when it comes to their rehabilitation progress?"

As the interviewed ILLAs did not have an active relationship with their HCP, the topic of goal setting was not discussed in detail. Any guidelines or recommendations that were set out for them during initial rehabilitation were simplistic, mirroring how the HCPs discussed goals with their clients. One ILLA for instance was told to try doing several exercises to lose weight. It was inquired whether the feedback received from their HCP was framed as positive or constructive, and whether they felt that the type of feedback motivated them. On reflection this question should have been targeted at the HCPs, as the effect of the type of feedback may not have been made readily apparent to the ILLA. Regardless, 5 of the interviewed ILLAs had only experienced positive feedback, while one ILLA received constructive feedback regarding their balance. Similar to their general experience with the rehabilitative process, all interviewed ILLAs found the way in which feedback was framed and handled as informative and motivating.

3.2.2.4 The Lower Limb Amputee Activity Profile

All ILLAs were assessed for their K-Levels of mobility, first by asking directly whether they knew or by summarizing the descriptions of the K-Levels (Medicare, n.d.) (via (Balk et al., 2018)), and asking which K-Level they thought was most relatable. As this was not carried out by a professional, these rankings should not be taken as their official levels of mobility. 1 ILLA considered themselves at K4, 4 at K3 (including the bilateral amputee; ILLA3) and 1 at K2. One of the K3 level participants, ILLA5, would have been at K4 levels in the past but due to their age had experienced natural degradation in mobility. ILLA1 likewise was at K3 levels of mobility in the past, but due to a recent stump injury had been reduced to K2.

The physical activity profile of an ILLA was obtained by asking about what activities they currently carry out, as well as any activities they desired to carry out in a realistic scenario. From the other perspective, the HCPs were asked about the kinds of activities their clients would carry out on a regular basis, additionally inquiring whether these activities would be carried out with or without a prosthesis. The frequency of activities that were included in responses are illustrated in the word cloud diagram in Fig. 3.1.

Figure 3.1 Word cloud of activities carried out by ILLAs

The frequency of the activity mentioned in dialogue corresponds to the size of the word, so for example "walking" was the most frequently mentioned activity. The frequency of walking in the word cloud was partially due to asking if participants went out on walks, as walking may not be considered by some to not be vigorous enough to be a physical activity. “Slopes” and “stairs” were also included if there were mentions of hill walking or moving about floors in their home. Swimming also appeared very frequently, but this is primarily due to asking HCPs about activities carried out without a prosthesis, for which swimming was a popular answer. Only 1 ILLA (ILLA5) mentioned they performed swimming, and another ILLA (ILLA1) desired to go swimming once their physical health had improved. Yoga and Pilates were often mentioned regarding maintaining stump mobility. For certain activities, the ILLAs were pressed to give more details about the activity regarding the frequency in which they would perform it and the difficulty of subsequent performance. There were a variety of descriptions relating to the frequency and difficulty of walking; some were happy to go out on walks every day, while others only walked if they needed to. Some ILLAs could go up slopes without difficulty, while others tended to avoid them where possible. Walking up a slope was generally considered more difficult to do than downslope, citing the design of their prosthesis as a limiting factor on their mobility during these movements. All ILLAs could climb upstairs without any considerable issue. On the other hand, downstairs movement was described as dangerous by a few of the ILLAs because of the higher probability of falling and greater risk of severe injury. The design of the stairs can also have an impact on whether the ILLA decides to traverse them: narrow stairs with irregular step height would be avoided in favour of wider steps with consistent step height where possible. When comparing between slopes and stairs, the general sentiment was that there is no clear-cut answer which was harder to traverse. Many factors such as step width, slope angle and uneven terrain can impact on how difficult it is to traverse them. The ILLAs were further quizzed on how often they wore their prosthesis, and whether they used mobility aids (for example, a wheelchair) in place of traversing with the prosthesis. All ILLAs wore their prosthesis as much as they could when traversing outside, and avoided using mobility aids unless completely necessary. All ILLAs except ILLA1 were also comfortable wearing the prosthesis at home; ILLA1 explained they felt more comfortable not wearing the prosthesis when sitting about. The others would only doff the prosthesis for certain situations like bathing, sleeping or if they felt pain in their skin around the area of the prosthesis.

A common sentiment among the ILLAs was that, while some wanted to get back to the fitness levels they were at prior to their amputation, none were interested in exceeding beyond those levels. The interviewed ILLAs, ILLA1 and ILLA2, both independently felt dissatisfied at their current activity levels and explained that their comorbidities and the pain associated with these comorbidities were the primary factor limiting them from being more active. ILLA3 and ILLA4, who had both had their amputations within the last 5 years, also expressed a genuine desire to improve their fitness levels. ILLA3, who had bilateral amputation, had expressed that they wanted to start running 10k races again. On the other hand, not all interviewed ILLAs seemed interested in improving their fitness. ILLA5, who was retired, did not want, or simply could not get back to their prior fitness levels and seemed at peace with the reality of their situation. To compensate, they took up hobbies that are less stressful to the legs and hips by going cycling or swimming. ILLA6, who was also retired, seemed content with their hobbies (walking and playing darts) and expressly mentioned they were satisfied with their current lifestyle.

3.2.3 Activity Monitoring

In the final section of the interviews, the topic of the questions deviated between the two groups of participants. HCPs were asked about what they would ideally want from an activity monitoring system. The ILLAs on the other hand were asked if they carried out some activity monitoring for their own personal gain. They were also gauged for their technological literacy to determine whether an activity monitoring system that was accessible to the ILLA demographic could effectively be utilized and understood by them, for example through an app or website.

HCPs were first asked what was important for wearable activity monitoring sensors in terms of its physical factor. Most HCPs did not have a particular preference, but two common qualities that were sought were that it should be tamper-proof (i.e a patient could not “shake” the device to simulate activity) and that it should be easily securable yet also easily accessible for recharging. Within the confines of a wearable sensor, the HCPs were also asked what kinds of physical activity would be useful to monitor. This question was similar, but slightly different to the question in the previous section where participants were asked about the kinds of activities they carry out. Some of the included activities, like amputee football, golfing, and darts, would not be possible to measure without a multitude of wearable sensors (Nguyen et al., 2015), and the sheer diversity of these activities relative to the number of ILLAs actively participating in them would not make them practical to monitor within the confines of the thesis. Thus, this question gives a more realistic take on the practical useful outcome measures that an HCP could utilise to monitor their client’s activity. The results of this question are visually represented in another word cloud diagram in Fig. 3.2.

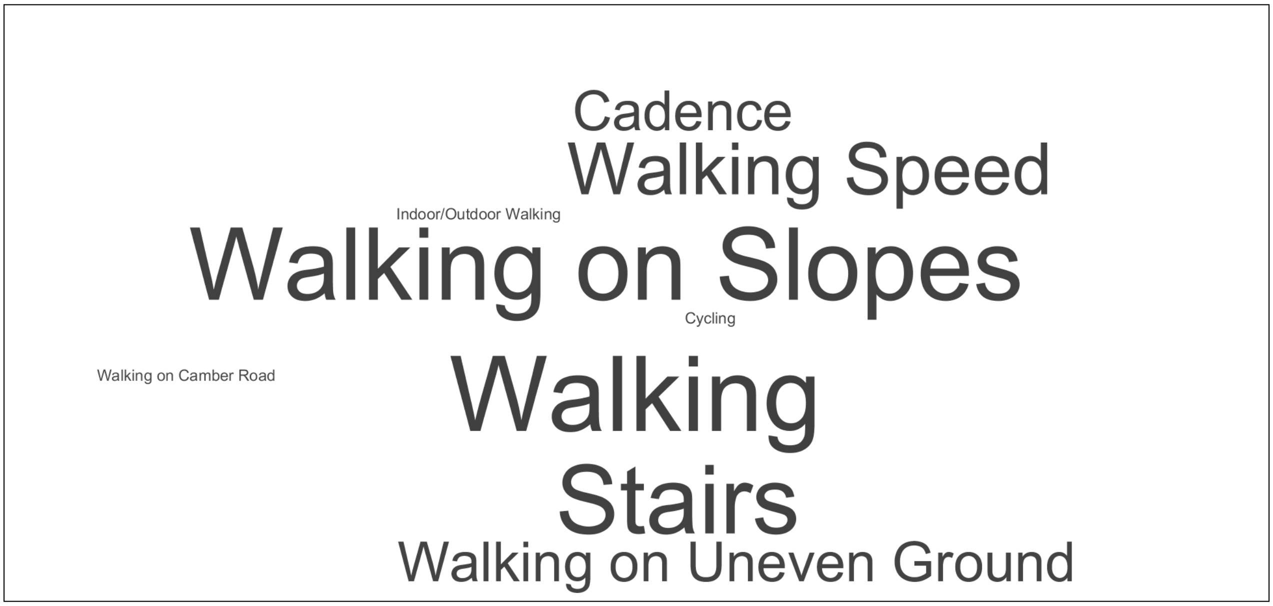


Figure 3.2 Word cloud of activities that are desirable as an outcome measure for the purposes of assessing ILLA patient physical health

Step count of slopes and stairs were some of the most desired outcome measures. Other desirable measurements were the ability to assess how long an ILLA wore their prosthesis (the "wear time"), how fast the ILLA walked with a prosthesis and for how long could they maintain that speed, how much time they were spending indoors as opposed to outdoors, and whether they could walk on varied terrain like uneven grounds or over the camber of a road. HCP4 gave valuable insight regarding the outcome measures that they could utilize in an activity monitoring system:

> *"It is important that an amputee is capable of taking their prosthesis for a walk upstairs or downstairs, on uneven ground, and outside when it is raining, and the ground is slippery. But the amputee might never bother going outside when it is raining, so they just never run into any issues with their prosthesis and will not rate their prosthesis experience as problematic, and so the problem is that*

> *they just do not do any activity beyond basic home ambulation and are not giving a proper assessment of their prosthesis. Another issue is that some amputees can develop workarounds to doing certain tasks, for example going downhill they might move sideways. While they are not walking correctly, because they are doing the task, they might see themselves as not having issues with it, and we have no way of knowing whether they're doing that or not."*

By being able to provide a method of recognizing a physical activity, a wearable sensor can address these concerns put forward by HCP4. Notably, if an ILLA traversed hills in an unusual manner like the quote describes, they would likely not be recognized by an activity monitoring system. This could then lead to the HCP and their client discussing how they are moving during a downhill segment.

An important component of an activity monitoring system is how the information is related back to the client, in this case the ILLA. While it was made clear that face-to-face feedback was the preferred form of communication for the HCP, some conceded that being able to relay simple and short messages through a mobile app or text could help supplement their existing systems. Likewise, the majority of the ILLAs were open to the idea of receiving feedback through texts or an app. HCPs were further asked, if the activity monitoring system were providing feedback remotely, what would be the recommended period between feedback sessions. There were lots of differing opinions, but the apparent general consensus was that any kind of feedback should be spaced at least a week apart, with some recommending months or even annual periods of feedback.

The interviewed ILLAs were asked how comfortable they were with operating certain kinds of technology, the specific technologies mentioned were laptops, personal computers, phones, tablets, smartwatches, and fitness trackers. Most were comfortable with operating the devices that were described, which aligns with research that has shown an increased uptake in technology with older adults (Mitzner et al., 2019). Only one amputee was uncomfortable with a using a smartphone or tablet and another was uncomfortable with using a personal computer, while the remainder were very comfortable using both devices. Two ILLAs used a fitness tracker (in the form of a smart watch) to monitor their fitness. In contrast, when asked whether their patients could use PCs/Laptops and other kinds of technological devices, the HCPs were either unsure or believed most of their clients would not be able to operate them well. Regardless, most HCPs conceded that there were an increasing number of their elderly patients who could operate a mobile phone.

3.3 Recommendations and Suggestions

3.3.1 Intended Audience and Usage for an Activity Monitoring System

From the discussions with both parties, it seemed clear that a novel activity monitoring system would be best suited at targeting the HCP demographic as the primary user of the system, through which they would be able to monitor their clients. Specifically, the system would benefit physiotherapists whose primary interest will be in keeping their patients physically active and healthy and may be especially useful for monitoring older dysvascular-origin amputees who may not have a strong drive to improve on their fitness compared to younger and traumatic-origin amputees. The information relayed by the physiotherapists could help inform the prosthetists on certain clinical decisions, for example, if a certain group of patients with one prosthetic fitting had significantly lower activity levels than another group with a different prosthetic fitting, that could indicate an issue with that prosthetic component. All the HCPs concur to an extent that they do not have a way of truly knowing how active their patients are, and some of them directly mention how the use of activity monitoring outcomes could be beneficial to them. HCP3 stated:

"If they have (physical activity) data to back up their claims that would be very

helpful as we would be able to see if they're lying or not without having to be accusatory."

One of the physiotherapists, HCP1, further added:

> *"With this activity monitoring system, I like the idea of being able to view the data and being able to do something with it. So, for example if I can see the person is being inactive, I can call them in and do something about it or refer them to a gym where they can do that activity. If a patient were claiming to be very physically active and were ready to be discharged from myself, I could view their data and see whether they were actually doing it."*

From this interview project, it is clear there is a genuine interest from HCPs to implement an activity monitoring system into their line of work. Although there is some potential in the future to provide an outlet through which the patient could additionally interact with the data (e.g through a mobile or web application), a novel activity monitoring system would be unlikely to be utilised by the average patient; though the majority of the interviewed ILLAs had an active lifestyle as described in Section 3.2.2.4, and claimed no issue with handling smart devices as described at the end of Section 3.2.3, they further went on to show a lack of interest in having their activity monitored, and the two participants who were interested (ILLA3 and ILLA4) already monitored their activity using a smartwatch. HCP8 succinctly described the dilemma of trying to create an activity monitoring system that targeted patients as the end-user; an app directly designed for an ILLA would not be very useful for the general population of ILLAs, who are primarily elderly seniors with dysvascular disease background, and will be unlikely to have a desire to track their own activity. On the other hand, the younger ILLAs who might have an interest in monitoring their health, would probably already be using an app or a fitness tracker. Referring to Section 3.2.2.1, it was found that all interviewed ILLAs did not have an active relationship with their healthcare professionals. This is a significant finding because it means that, once the ILLA has been discharged from their initial rehabilitation setting (e.g a hospital or a clinic), their contact frequency reduces significantly over time. If the activity monitoring system uses a wearable device, these devices may end up being forgotten or lost. Thus, taking these arguments into consideration, the framework should not be designed to target experienced amputees due to a combination of a predicted lack of interest and practical limitations of communicating monitoring data from patient to HCP (or vice versa) when interactions between the two parties exceed several months.

A key finding from Section 3.2.2.1 was that most ILLAs did not feel like their rehabilitation sessions could have been improved. When asked how their rehabilitation sessions in the past could have been improved, none of the ILLAs could come up with any ideas. ILLA2 added that:

> *"to be honest I don't like the idea of telling (their physiotherapist) how to do their job better."*

If these findings are generalizable to the sentiments of the ILLA population, then ILLAs could see the methodology set out by their HCPs as being integral to their own rehabilitation progress, and so if a HCP were to suggest to a client that they could monitor their activity with an activity monitoring system, there is a reasonable chance the ILLA will abide by these suggestions if they trust their HCP. Taking the findings from Section 3.2.2.1 into account, the ideal use for an activity monitoring system appears to be when ILLAs are in the earliest stages of their rehabilitation. At this stage, the healthcare professional maintains a close relationship with their ILLA, meeting with them sometimes several times in a week. The activity monitoring system could be introduced at this stage, and as Section 3.2.2.1 infers, most ILLAs will see this as an essential part of their rehabilitation. This could also allow the HCP and their ILLA to maintain a more proactive relationship once the ILLA has been discharged from the hospital or rehabilitation centre, as they would have more interactions through discussing the findings from the system. If the system were targeted at discharged ILLAs who have spent months or even years without interacting with their HCP, the reception would likely be much more negative as it would be forcing them to maintain or rekindle a correspondence, which they may not want to have.

3.3.2 Key Activity Monitoring Outcomes

The key activity monitoring outcomes from the word cloud diagram in Fig. 3.2 were the outcomes that could be optimistically collected from a wearable sensor. While ILLAs can choose to perform activities that do not require a prosthetic leg, such as wheelchair sports, given that prosthetists were one of the stakeholders of the research outcomes, the author decided to focus on monitoring activities where use of a prosthesis would generally be required. As the research progressed, there were limitations rooted in the chosen activity monitoring sensor (the process of selecting an appropriate sensor is discussed in Chapter 4.3.3) as well as some practical study limitations that prevented the monitoring of all outcomes listed in the diagram. This is discussed further in Chapter 5.9.2. Future studies with different monitoring sensors and less restrictive ethical limitations could theoretically measure more - if not all - of the activity outcomes shown in Fig. 3.2. Additionally, some activities from the first word cloud (Fig. 3.1) could also be collected in the future – specifically, those with repetitive and homogeneous motions such as operating a rowing machine or doing yoga exercises could theoretically be programmed to be recognizable in an activity monitoring system. The collection of such activities however is outside the scope of the thesis.

3.3.3 Comparisons with Similar Research

The findings of this study are supported by other investigations in the field of research, who have employed larger scale focus groups with higher numbers of participants and more diverse healthcare professional backgrounds. Klute et al. (2009) conducted focus groups with ILLAs, clinicians, researchers, and prosthetic device manufacturers to understand the primary needs of ILLAs. One of the key findings was that there was a need to implement "remote monitoring systems" to improve on the standards of care. Their methodology in terms of how the focus group was conducted was also notable; they held an initial session which consisted of all stakeholders in a single focus group, where they identified the desires from each group relating to their technical goals, how each group communicated with each other, how they would learn from each other, and how the healthcare of the ILLAs should be handled. They then held separate sessions with each group individually where they would discuss in a workshop-type format how they could come up with some solutions to their key needs. In this research, if it were possible to combine the two groups in a single session, it is likely the discussion sessions would have been more productive. To give a specific example, it would have been interesting for the prosthetists and physiotherapists to learn why the majority of the interviewed ILLAs no longer received physical therapy or contact from their own HCPs, which could have led to some discussion on how their relationship could be maintained for a longer period. A noteworthy point of discussion in Klute et al. (2009)'s work is that when discussing a remote sensing platform, they include a shared number of concerns for its implementation, that being the ethical concerns of sharing data with HCPs and how the system would be automated regarding data being sent to the healthcare providers. These ethical concerns are addressed in more detail in Chapter 9. They also expressed similar concerns about the physical form of the wearable device, citing the reliability and power requirements as two particularly important issues. These issues are discussed further when describing the wearable sensor selection process in Chapter 5. Dekker et al. (2018) conducted a focus group study on the needs of ILLAs via focus group sessions held with HCPs, and likewise concluded that there should be improved remote monitoring of ILLAs, especially younger dysvascular-origin amputees, in the early stages of rehabilitation. Unfortunately, neither Klute et al. (2009) or Dekker et al. (2018), nor any other focus group study to the author's knowledge have specifically targeted interviewing ILLAs in the early stages of post-amputation rehabilitation regarding which specific physical activities are carried out or are desired to be carried out by the demographic. Watters and Deans (2015) discussed activity and barriers to physical activity in an ILLA focus group but did not specify the experience of the ILLA group. Fogelberg et al. (2016) discussed physical activities carried out with prosthetic foot users but excluded potential participators if they had K1 levels of mobility. Van Den Akker et al. (2020) interviewed wheelchair users, of whom several of which had lower limb amputation. However, none of the interviewed ILLAs mentioned a desire to use a prosthesis or other upright walking aids. Whether this was due to a lack of functional or voluntary capability was unclear, but as the primary focus of the study was on people with spinal cord injuries, this would indicate it is the former. All ILLA volunteers in the focus group of Day, Wadey, and Strike (2019), who discussed the daily challenges of performing physical activity had at least one year of experience post-amputation.

By widening the scope of research to include focus group studies of physically vulnerable populations (including ILLAs), there have been calls by these studies for the adoption of activity monitoring systems in clinical care; Wu et al. (2019) used focus group sessions to assess the practical feasibility of activity monitoring patients with Gastric cancer by tracking their activity with a mobile health application and monitoring to see if they utilised the application. Although the overall period of monitoring was short (1 month maximum), 86% of all participants used the device for all 4 weeks, indicating viability of an activity monitoring systems for short to mid-term durations. Kononova et al. (2019) likewise found in their activity monitoring focus groups that there was potential to promote physical activity in elderly populations and explained that providing social support and motivation from the HCP was key to ensuring that engagement would be maintained in the long term.

3.3.4 Study Limitations

The lack of volunteers for both for the ILLA and HCP demographics limited the generalizability of the findings. In the future, follow-up focus group studies should discuss the conceptual activity monitoring system developed in this thesis to confirm whether there is a genuine need and desire for the system in a clinical setting. This will require targeting much greater volumes of volunteers from both demographics. As mentioned, there was some location bias in the ILLA demographic, which could have been mitigated by increasing the scope of recruitment outside of Glasgow. By providing reward incentives, it may have been possible to entice lesser physically active ILLAs and gain a more diverse outlook on experiences. There were considerably fewer recruitment limitations in the HCP demographic, however it would have been beneficial to have recruited other stakeholders whose profession closely relates to this research, such as occupational therapists, surgeons, gait analysts and other researchers (Dekker et al., 2018).

A key issue to address regarding the focus group was that all the ILLA participants were already involved in volunteering at the NCPO to assist with student prosthetist training sessions, and the focus group was carried out in conjunction with one of these sessions. This highlights a strong case of volunteering bias; these participants were willing to take part in hour-long volunteering sessions, which implies that they are all physically fit to a large degree relative to their condition and were all active community participators. Had participants included ILLAs in the early stages of rehabilitation or those with adverse comorbidities (for example, obesity or Chronic Obstructive Pulmonary Disease), the desire to improve or even maintain a regular level of fitness may not have been as prevalent. Similarly, all interviewed ILLAs preferred ambulation with prosthesis over the use of walking aids or wheelchairs, but this may not be case for ILLAs in the early stages of rehabilitation, who may not have as much confidence with their prosthesis. Being part of a series of volunteering sessions meant that the volunteers were, at the very least, all acquainted with each other, and this would have impacted positively on the focus group's social dynamics: it is likely the case that the volunteers felt more inclined to share information with people they were familiar with.

The interview and focus group sessions were also generally limited by the inexperience of the interviewer; it was their first time conducting these sessions in a formal context. In the focus group session, they had likely missed out on some key facial expressions and body language which could have given more insight into their true feelings. Future interview sessions should use an assistant to note down any distinct facial expressions or use a video recording so that visual information could be retroactively reviewed. The focus group session and some of the interviews were also held at the end of the working day, and so the interviewer occasionally felt pressure to press through the questions without expanding on some answers so that their participants were not fatigued. Conducting these sessions during the day, while providing refreshments where possible, would have helped improve the productivity of these sessions.

3.3.5 Summary & Implications of the Interviews

From a qualitative analysis of the series of interviews that were carried out, the following interpretations relating to the research were made:

- Physical activity monitoring is not currently employed in clinical practices, at least in terms of application of clinical evaluation of ILLAs.
  - The barriers to implementation include lack of resources, failed attempts in the past (e.g through mail-out questionnaires) and concerns of validity of measuring devices
- All interviewed HCPs expressed some interest in capability of employing physical activity monitoring in their profession
- The interviewed ILLAs did not keep in touch with their HCPs outside of emergencies, indicating that physical activity monitoring for the purpose of clinical evaluation only has practical use in the early stages of the ILLA's prosthetic rehabilitation, likely in the prosthetic training phase.

- Though some of the interviewed ILLAs used some form of activity tracking (e.g through a smartwatch), they did so after their prosthetic training was complete and no longer kept in regular contact with their HCPs.

- The interviewed ILLAs were positive about their experiences in the early stages of prosthetic rehabilitation, understanding that the treatment and advice given by their HCP was vital in accelerating the rehabilitative process, thereby indicating that employment of physical activity monitoring during the prosthetic training stage would be met positively

- From a discussion of the activities with ILLAs and HCP perspectives, a collection of physical activities performed by ILLAs were established. Through further discussion of activities that would be beneficial to evaluate in an activity monitoring system for HCPs, a list of clinically relevant outcome measures were established.

  - These outcome measures were: stepping activity on slopes, stairs, camber (of a road), uneven terrain, and indoor/outdoor stepping, as well as prosthetic wear, cadence variations, walking speed variations and going cycling.

Modern day healthcare of ILLAs still has a long way to go before an activity monitoring system can be implemented into current healthcare plans, which has been evidenced in this Chapter. The literature review by Chadwell et al. (2020) of monitoring technology for lower and upper limb prosthetic use supports this argument; in many countries, remote prosthetic monitoring technology is often expensive, has low availability and limited battery life. It is beyond the scope of the thesis to tackle the socioeconomic barriers that are throttling its implementation. Instead, the research will focus on how the clinically relevant outcome measures identified in this review can be recognized through sensory detection, while also aiming to tackle the barriers to prosthetic monitoring identified in Chadwell et al. (2020). These clinical outcome measures will them form the basis of a clinical activity monitoring system, the framework of which is proposed in Chapter 9 following an evaluation of the reliability of the outcome measurements in Chapters 5 - 8.

With a list of clinical outcome measures of physical activity established, the next stage was to establish a process of measuring these activities. From Chapter 2, it was determined that subjective evaluation of physical activity, i.e through questionnaires, is an unreliable method. Affordable commercial fitness tracking devices like the FitBit® offer limited capacity to monitor more detailed stepping activity, like measuring going up flights of stairs (Fitbit, n.d.[a]), however these measurements have not been validated for their accuracy, nor can they track downstairs movement. State-of-the-art fitness trackers like the Polar Grit X (Polar Electro, Zug, Switzerland) can monitor changes in altitude while walking with reportedly good accuracy (Polar-Electro, n.d.; DC-Rainmaker, n.d.), but relies on Global Positioning System (GPS) co-ordinates (thereby cannot track indoor activity) and its expensive MSRP would inhibit implementation in clinical settings. These findings show that no single device can measure the desired activity monitoring outcomes purely by extracting information processed from the device. Therefore, it is essential to develop a new algorithm from a sensory device that can acquire these outcomes. It would be extremely difficult to create a purely logic-based algorithm to capture these different kinds of activities; it would require a vast understanding of how the sensory parameters change with the different types of outcomes, and would also need to be generalizable to ILLAs with varying levels of amputation and mobility. The most practical solution to creating an algorithm is to employ artificial intelligence. Artificial intelligence, in the context of detecting activities carried out by people (including ILLAs) is known as Human Activity Recognition (HAR). HAR is the process of creating an automated system that can detect activities carried out by an individual or group of individuals. A review of modern advancements in HAR is discussed in Chapter 4, and through the review a novel HAR study of ILLA activity monitoring is proposed that captures the clinical activity monitoring outcomes discussed in this chapter.

3.4 Chapter Conclusion

This chapter has demonstrated the key qualities that would be sought in an activity monitoring system for ILLAs. Primarily, the important activity monitoring outcome measurements have been identified and the stage in the rehabilitation process for which this system would be at its most effective was also deduced. To capture the identified activities in a reliable manner, it was determined that no current off-the-shelf activity monitoring system was validated to, or capable of, capturing the identified measuring outcomes, and a new activity monitoring system would need to be designed. This was achieved through the application of machine learning and human activity recognition, which is discussed more in the next chapter.

Chapter Four

Building Towards a Novel Solution: A Review of Human Activity Recognition

4.1 Introduction

In Chapter 3's concluding pages, it was established that machine learning would be required in order to reliably recognize the desirable activity monitoring outcomes for a clinical activity monitoring system for ILLAs. The aims of this chapter are twofold. Primarily, the chapter will give a detailed review of how machine learning is applied for the purposes of HAR. Secondly, the chapter will summarize the review of the choices that can be made in machine learning and based upon the established literature and the findings of the interviews from Chapter 3, model a novel HAR study for ILLAs. Should the HAR system be validated with high accuracy, precision, and recall, the work could form the basis of a clinical activity monitoring system and be implemented into a framework to facilitate the prosthetic rehabilitation of newly-amputated individuals.

4.1.1 Establishing the Review scope

Machine learning is an adaptable and versatile tools in modern day data sciences. It is a branch of artificial intelligence, where a computing system 'learns' how to make predictions on future events by being 'trained' by data from past events. A common example of application of machine learning in ILLA rehabilitation is fall detection, which can be a serious risk for elderly ILLAs (Daines et al., 2021; Engenheiro et al., 2020; Shawen et al., 2017). However as an ILLA falling is not conducive to the rehabilitation of the ILLA, it is not within the scope of this review to discuss fall detection in great detail, although some of the included papers in this review have analysed fall detection as a secondary outcome. Additionally, collecting data on fallers would not have been practical given the limitations of the data collection process as discussed in Section 4.3.2.

Recognition of HAR can be divided into two broad categories: high-level and low-level activities (Khowaja, Yahya, and Lee, 2020). A "low-level" activity is a simple action, usually repetitive in nature and generally immutable - that is, changing how the activity is performed would change the definition of the activity. These include activities like walking, cycling, running and stretches. A "high-level" activity is generally comprised of a sequence of low-level activities, and unlike with low-level activities there are numerous ways in which a high-level activity can be performed. Such activities include cooking, cleaning, brushing teeth and sports activities. As the activity monitoring outcomes identified in Chapter 3 are exclusively low-level activities, they are the focus of the review, although some of the articles included in this review have partially incorporated high-level activity recognition.

In regard to the sensory deployment used in HAR, the scope of the review is limited exclusively to wearable sensors, meaning sensors that are attached or worn by the subject. The two other typical modes of sensory deployment are on-object sensors and video monitoring. On-object sensors refer to any sensors that are embedded into everyday objects, where interaction with said objects triggers detection of a physical activity. On-object sensors can contain accelerometers, thermometers, barometers, magnetometers, GPS and Radio-frequency Identification (RFID) readings. Relating to the clinical outcomes discussed in Chapter 3, on-object sensors would have been impractical to deploy regarding monitoring of walking activities. In a study by Logan et al. (2007) which used over 900 on-object sensors, it was found there were high maintenance costs and significant problems with data asynchrony, which would suggest that there are practical limitations to the number of sensors that can be used. In video monitoring HAR, images from camera feeds can use properties in pixel intensities to gather machine learning features and allow machine learning algorithms to process human activities, even in real time (Htike et al., 2014; Leo, D'Orazio, and Spagnolo, 2004; Lin et al., 2008; Robertson and Reid, 2006). Once again regarding clinical activity monitoring outcomes, the deployment of cameras for video monitoring of walking activities would not have been practical; in the context of clinical applications, deployment of the systems e.g for at-home monitoring would have serious ethical considerations for the privacy of the individual.

Finally, due to the general lack of HAR studies specific to the ILLA populations, the scope of the review is widened to include HAR studies of the general population as the target audience, as it should be considered that the machine learning techniques used will be fundamentally similar.

4.2 Part I: The Machine Learning Process for HAR

Activity data in the form of sensor information is initially acquired either first hand or through a pre-existing dataset. From the sensors data, a set list of properties, known as 'features', are calculated, which will then be fed to a machine learning classifier to establish a mapping function, which determines in a systematic fashion how an unknown point of data is associated with a 'class'; in this context, a class generally means an activity carried out by an individual. The overall performance of the classifier is assessed through validation, which typically involves splitting the data into 'training' and testing' partitions. The trained partition trains the machine learning classifier with 'known' data, as in the data whose labels or class membership are known. The trained classifier will then predict the class of the testing data, which can be compared to the ground-truth to assess whether it has made the correct prediction.

One of the main challenges of HAR is that each step in the classification process involves making choices which can greatly impact on the performance of the machine learning system. The first part of this chapter will look at these steps and the choices made in these steps in greater depth. It should be noted that the chapter does not discuss several design steps which are very common in the HAR process: namely, the data annotation process, pre-processing techniques, and segmentation of the data into discrete samples. Because the suitability of these processes will change depending on the activity monitoring set-up, they are discussed further in Chapter 5 following the establishment of core methodology of the activity monitoring study in this chapter.

4.2.1 Data Sources

4.2.1.1 Public Datasets

Data collection is, by and large, an expensive and time-consuming endeavour. By using publicly available datasets, the researcher can cut down on a considerable amount of time and money expended on research. Additionally, machine learning studies that use the same dataset can be conveniently compared for differences in machine learning techniques. These pre-existing datasets and the research papers associated with them provide a solid foundation on which future machine learning studies can compare, contrast, and improve. The UCI repository contains a wealth of human activity recognition datasets that have formed the basis of many machine learning investigations (UC-Irvine, n.d.).

Demrozi et al. (2020)'s survey on sensors used in HAR provides an up-to-date, comprehensive list of datasets used in HAR; 30 publically available, 112 under restricted access. To summarize their key findings: all datasets used inertial sensors. All datasets used an accelerometer, occasionally accompanied by a gyroscope and less frequently a magnetometer. The number of participants in HAR datasets varied from 563 participants in the WISDM v2 dataset to one participant in the Skoda dataset (Kwapisz, Weiss, and Moore, 2011; Zappi et al., 2007). Datasets observed an average of 12 activities. In the majority of datasets, the activities monitored included "low-level" activities. When research prioritizes the general applicability of a machine learning technique (for example, testing a novel classifier), datasets provide a simple and effective method of applying that technique in a wide variety of scenarios.

4.2.1.2 Data Collection

A standard method of obtaining first-hand data for HAR is controlled data collection. The participant will be placed in a constrained environment under supervision of the research team and instructed to carry out sequences of physical activities while their movements are captured by sensors. This approach gives the researcher the maximum amount of control over their environment and usually makes the data annotation process straightforward. Due to its consistency and reliability, many controlled data collection investigations have led to publicly available datasets; over 75% of datasets in Demrozi et al. (2020) were carried out in controlled environments. The drawback of this approach is that the collected data may not generalize to activities carried out in real-life conditions. In fact, machine learning algorithms that have been trained on lab-based datasets are at risk of performing poorly when applied to naturalistic conditions (Schrader et al., 2020).

In naturalistic data collection, the subject performs physical activities in unconstrained environments or free-living conditions (e.g their home, a park, community centres, gyms etc.) with or without direct observation from the researcher team. While this approach can be seen as the most realistic method of activity collection, naturalistic data collection comes with significant caveats. Manual annotation becomes a difficult task when activities can transition erratically and ambiguously. The ground truth must come either through the researcher's observation of activities, self-annotation methods (e.g an activity diary) or video recordings (Aloulou et al., 2017; Bao and Intille, 2004; Zhao et al., 2013). Each method of validation has its own flaws: observation of activities by the researcher is not practical during long term monitoring and scales poorly in large populations, self-annotation techniques can be unreliable, and video recording is limited by the number of cameras available, their battery life. Video recording further risks obfuscation of the camera lens (Woznowski et al., 2016). Due to their price, camera recording also scales poorly as a validation technique with large numbers of participants. Additionally, without observation and feedback from a researcher, the participant is at risk of being interrupted or distracted from the task at hand, reducing the volume of viable data.

4.2.2 Sensor Configuration

If the data is being collected first-hand, the next step in the decision process is to determine which sensors are needed, and where they should be placed. Even when the data is being obtained from a dataset, the information regarding the sensors play a key role in deciding the appropriate feature selection and classification techniques, as the type of sensors and their subsequent placements can have significant influence on the data.

Broadly speaking, a physical activity will produce some change in the monitoring sensor's values. This sensory data can then be passed to a processor, which will use algorithms to translate the change in sensory values to a detection of some form of physical activity. The advantage of adopting sensor-based monitoring is that quantitative data is inherently generated from the process, which means that the end user (e.g the clinician and/or their client) have a means of tracking physical activities over time. The evaluation also does not have to rely on the ILLA's memory or biases, which can easily influence outcomes in subjective based monitoring. This does not mean however, that sensor-based monitoring of physical activities is functionally perfect and come with their own set of unique limitations. This section will review the types of wearable sensors that are used in HAR studies.

4.2.2.1 Sensor Modality

All HAR studies use some combination of sensors in order to achieve their objective, therefore the choice of sensor is often the most important stage in the entire design process. Common types of sensors that are used in HAR are discussed in this section. In relation to wearable sensors, there is some additional discussion on the importance of their placement on the human body.

4.2.2.1.1 Mechanical Sensors Mechanical sensors detect some form of mechanical deformation (as generated by human motion) and translate the deformation into an electrical signal. Of all mechanical sensors (and wearable sensors in general), accelerometers are the most abundant method of measuring objective physical activity counts (Demrozi et al., 2020). When the human body produces movement in proximity to the accelerometer, this generates a force on the proof mass which generates a change in electrical signal through a key mechanism: usually a piezoelectric/piezoresistive effect or a change in capacitance, depending on the design.

Accelerometers are an extremely popular sensing modality for HAR applications. Khan, Lee, and Kim (2008) created an algorithm using only a single chest worn tri-axial accelerometer and was able to achieve near-perfect recognition accuracy (99% on all counts) to recognize walking, sitting, standing, and lying. In Chen and Xue (2015), an algorithm with a single smartphone sensor was able to distinguish 8 activities[1] with an average accuracy of 93.8%. When the accelerometer is the chosen sensor for HAR, some further considerations to the

[1] Activities monitored: falling, jumping, running, walking, step walking, walking quickly, downstairs, upstairs

specific make of sensor must be made; Analogue to Digital (ADC) resolution in the accelerometer for example, is often an overlooked aspect in accelerometer choice (Twomey et al., 2018). Higher ADC resolution could lead to improved HAR performance, but at the cost of dramatically reduced battery life which is unsuited for long-term monitoring. Khan et al. (2016) discusses how accelerometer frequency should be optimized for HAR. They argue that, although most human activities do not exceed 10Hz, sampling frequencies should exceed the Nyquist 20Hz regardless. They explain that this is due to the lack of recurrence in accelerometery signals which can make signals appear different at higher sampling frequencies. While this would indicate higher frequency sampling is superior, much like ADC resolution, there will be a trade-off in HAR performance and longevity of the sensing platform. Although the choice of sampling frequency is problem-dependent, Khan et al. (2016) demonstrates that the best frequency will likely range somewhere between 20 and 45Hz. Alongside accelerometers, gyroscopes and magnetometers are utilised as complementary sensors. A common term for the fusion of an accelerometer and a gyroscope is the Inertial Measurement Unit (IMU), which monitors changes in motion with six degrees of freedom. This can be further combined with a magnetometer to measure three additional degrees of freedom.

The function of a gyroscope is to measure the angular rotation rate around an axis, which can be further applied to estimate the pitch angle (Sabatini et al., 2005), though this is also achievable with just an accelerometer (Pedley, 2013). When combined with the accelerometer, the IMU can detect both linear and rotational information. Successful applications of an IMU in activity monitoring include detecting 19 different exercises[2] and detecting the transition between poses in a yoga movement, both with reliable accuracy (Altun and

[2]Activities monitored: sitting, standing, lying down on back and on right side,ascending and descending stairs, standing in an elevator still, moving around in an elevator, walking in a parking lot, walking on a treadmill with a speed of 4 km/hr (in flat and 15 $\circ$ inclined positions), running on a treadmill with a speed of 8 km/hr, exercising on a stepper, exercising on a cross trainer, cycling on an exercise bike in horizontal and vertical positions, rowing, jumping, playing basketball

Barshan, 2010; Omkar, Mour, and Das, 2009). Unfortunately, gyroscope drift is a common phenomenon in gyroscopic sensors generated from bias instabilities and angular random walk which, due to integration factors, can produce large reading errors over long periods of time (Beavers, 2017; Yi Wang and Yun Meng, 2017), though this can be corrected with readings from the accelerometer (Li et al., 2019). Magnetometers measure relative changes in magnetic fields. Their insensitivity to acceleration makes magnetometers suitable for measuring orientation in human activity monitoring (Lee et al., 2016). Combined with accelerometers and gyroscopes, this gives an IMU 9 degrees of freedom. In the activity monitoring research field, magnetic induction based wireless systems have been used to successfully recognize a wide range of activities, whilst reducing power consumption compared other wireless network systems (Golestani and Moghaddam, 2020). On the other hand, magnetometer readings can easily be interfered by external magnetic sources (Wu et al., 2018). Introducing a Kalman filter can help to mitigate this interference (Beravs et al., 2014), as well as correct for gyroscope drift (Zhu and Zhou, 2004), however the technique becomes ineffective with prolonged magnetic interference and incurs additional financial, power and processing costs (Frick, 2015).

4.2.2.1.2 Physiological Sensors Physiological sensors use signals emanating from the human body, such as heart rate, respiratory circulation, or Electroencephalography signals to determine if the individual has undergone physical activity. For example, the SenseWear® Armband (BodyMedia, PA, USA) is a physiological sensor that measures the energy expenditure of the user from a multisensory array based on tri-axial accelerations, temperature, skin flux and galvanic skin response (English et al., 2016). Electromyography (EMG) signals have been used in studies to accurately classify precise human movements (Biagetti et al., 2018; Liu and Schultz, 2019; Zhang, Ling, and Li, 2019). Physiological sensors are used less as the primary measurement of activity compared to accelerometer-based signals and are more often used as a supplement (Chen et al., 2012; Qi et al., 2018). When EMG signals are utilized, they either have poor recognition performance or do not improve accuracy when fused with accelerometer data (Abdallah, Abdulsadig, and Amien, 2019; Biagetti et al., 2018). Heart rate signals likewise appear to have negligible impact on HAR performance when fused with accelerometer data (Balli, Sağbaş, and Peker, 2019).

4.2.2.1.3 Environmental Sensors Environmental sensors typically do not measure human activity directly, but rather measure some change in the environmental property that is generated by a human activity and uses the change in property to infer the action carried out by the person. Broadly speaking, temperature, humidity and barometric pressure are the main environmental sensors deployed in HAR (Demrozi et al., 2020), however their applicability only extends to niche activity detections. In Muhammad Masum et al. (2018), the only purpose of including temperature and humidity sensors alongside inertial sensors was to be able to detect the difference between sitting and sitting in a toilet. In De et al. (2015), "opening fridge doors" could not be recognized as the temperature sensor was too slow to respond to the change in ambient temperature. De et al. (2015) further goes on to show that barometer sensors were only useful as binary detection devices (e.g sitting on couch, sitting on floor). Studies have rarely investigated how environmental sensor data impacted on machine learning performance. In Kanjo, Younis, and Ang (2019), it was found that data trained only on these signals had the poorest activity recognition performance, with the maximum average recognition achieved at 60%. It is evident from Vaizman, Ellis, and Lanckriet (2017)'s study that environmental sensors do not necessarily improve activity recognition rates, but rather can assist with the annotation process. In their study, they used audio and GPS signals to help decipher the contextual environment where a PA was being performed, which in turn made the annotation process less error prone. Further evidence is shown in Md et al. (2016), which used heart rate and barometric sensors to "context filter" activities primarily measured through accelerometer and gyroscope sensors. For example, walking movement can be derived from rapid changes in accelerometer readings, then by further applying barometric readings to identify changes in altitude, this can help to contextually distinguish between level walking, ascending and descending stairs.

4.2.2.2 Types of Wearable Sensors for HAR

Wearable sensors refer to any sensory device that is attached to or worn by humans while performing physical activity. As such, all wearable sensors require compliance from the user to ensure that the device is worn and placed correctly in order to obtain accurate measurements.

4.2.2.2.1 Pedometers Pedometers were one of the first devices used to quantify physical activity in humans. Originally envisioned by Leonardo da Vinci, the first pedometers were said to be created in the late 18th century (McManus, 2015). These first pedometers were mechanically operated and relied on a pendulum which swung with the motion of forward gait. When triggered by the pendulum, a type of lever called an escapement would interact with a gear wheel inside a clock which would increment the time spent walking. Nowadays, most modern pedometers are entirely electronic and use an accelerometer to facilitate the step counting process through algorithmic detection. Step counting is one of the simplest ways of quantifying physical activity and is also a good indicator of health; a study carried out by Schmidt et al. (2009) shown that a high step count was associated with lower risk of adverse cardiometabolic effects. Step count can be derived into other clinically useful measurements, for example "cadence" indicates how many steps have been taken over a predefined epoch. Differences in cadences can give an indication of what the user is doing at the time; low cadence indicates incidental stepping, while high cadence indicates purposeful walking and thus a good indication of moderate physical activity. A clinical application of utilising cadence measurements is to equate the cadence to walking speed to determine older adults' mortality risk (Studenski, 2011). An issue that plagues even modern pedometer devices is their accuracy; dedicated step-tracking devices like the Omron™ have been found to consistently underreport step count even under controlled laboratory conditions, like walking on a treadmill (Husted and Llewellyn, 2017), thus only give an estimation and not an exact measurement of physical activity.

4.2.2.2.2 Smart Devices An increasingly popular method of measuring activity is by using accelerometers or IMUs embedded in smartphones. This method of activity monitoring is particularly advantageous from a practical and financial perspective due to the prevalence and popularity of smartphones in the modern era. This in turn reduces research and development costs as developers can focus entirely on their software implementation. Albert et al. (2013) for instance created a mobile application which can monitor activity levels in ILLAs and distinguish between inactivity, swaying, slow and fast walking. However, while many people own smartphones, the elderly population, which represent a sizable proportion of the ILLA population (Davie-Smith, Hebenton, and Scott, 2018), can be slow on technology uptake, and can have difficulty operating mobile applications that track their fitness (Gordon et al., 2019). Consistency and user compliance are also significant issues; accelerations captured by the mobile phone will be different depending on where the phone is placed (in a pocket, hand- bag, carried by hand etc.). Wrist-worn smartwatches and fitness trackers such as the Apple Watch®(Apple, CA, USA) and Fitbit® can mitigate this issue by having a consistent orientation for acceleration measurement, making them an increasingly popular way to track activity for ILLAs (Smith, Guerra, and Burkholder, 2019). Nonetheless, user compliance can still be a major factor; if the user takes off the device for charging and forgets to wear the device, a significant amount of physical activity data can be lost.

4.2.2.2.3 Insoles One method of increasing user compliance with activity monitoring is by directly embedding sensors into the footwear that the user uses for walking. A common location for embedding is the insole of the footwear, which has led to the development of 'smart insoles'. Common sensor modalities found in smart insoles include accelerometers, pressure sensors and thermometers, which can be utilised to measure a diverse range of physical activity-based measurements, and are even capable of physiological-based tracking measurements like heart rate and muscle activity (Subramaniam et al., 2022). Activity monitoring studies that have employed smart insoles have had consistently high accuracy in step counting, measuring between 94.8 – 100% across a variety of studies (Ngueleu et al., 2019), while also being found to be cheaper to manufacture compared to other popular wearable activity monitoring devices like the Fitbit® (Lopez-Meyer, Fulk, and Sazonov, 2011). Specific to potential application of lower limb amputee evaluation, smart insoles are capable of accurately measuring cadence and temporal gait characteristics in post-stroke patients, indicating applicability in accurately measuring physical activity for those with asymmetrical gait (Lopez-Meyer, Fulk, and Sazonov, 2011). Though no studies were identified that tracked amputee physical activity using smart insoles, they have shown capability of reliably monitoring gait asymmetry in ILLAs in pilot study (Loiret et al., 2019), and by considering their performance from Ngueleu et al. (2019), this would suggest they are still well suited for that application.

While the literature shows strong support for the implementation of smart insole sensors for activity monitoring, they are not without some drawbacks. Battery life may be of concern; if the insole sensor does not send battery information to the user e.g through a smartphone app, the user will not be aware of how many remaining hours of operation the insole sensor has. Additionally, unless the insole sensor is custom fabricated for a specific set of footwear, the insole risks slipping on the foot during walking (Lee, Hong, and Oh, 2018).

4.2.2.3 Embedded Sensors in Prosthetics

State-of-the-art prosthetic components have the ability to perform activity monitoring using embedded multi-sensor arrays, including accelerometers and goniometers. Research by Maqbool et al. (2015) has shown that implementation of gyroscopes and footswitches can be used for real time recognition of gait events (heel strike and toe-off) under flat and sloping conditions, which could then be further implemented into an alteration of prosthetic properties – for instance increasing friction at the prosthetic knee joint during downhill descent to prevent high thigh flexion angles and causing a loss of balance. The Orion3™(Blatchford, Basingstoke, UK) has an inbuilt algorithm that can accurately monitor step count during dynamic forms of movement like stair and slope ambulation (Sykes et al., 2018). The embedded microprocessor in modern iterations of the C-Leg™can measure and project step count to smart devices via a companion app (Ottobock, Duderstadt, Germany), allowing remote monitoring of activity (OttoBock, n.d.). Similar to smart insoles, the advantage of these kinds of devices is that they facilitate compliance; if the user is wearing their prosthesis, activity monitoring information can be automatically obtained without demanding the user having to wear additional sensors. Likewise, this also provides an inherent method of analysing prosthetic wear time, as a total lack of movement/physical activity indicates the user has not worn their prosthesis. However, while compliance is facilitated, it is not guaranteed when the sensing unit needs to be recharged in order to function. Another drawback is that the information currently gathered by these devices often give very limited information relating to physical activity. Referring back to the C-Leg™ mobile app, users can only view their daily step count or total step count and cannot track their step data across the span of the day.

4.2.3 Number and Placement of Wearable Sensors

The number and location of the sensors in a wearable sensor HAR recording can play a significant role both in terms of what activities are measurable, and the resulting classifier performance. Typical body sensor placements include the head, upper limbs, chest, back, hip, waist, lower limbs, and feet, as well as body peripheral placements like bags or pockets where a smartphone would usually be located (Wang, Cang, and Yu, 2016). Huang et al. (2017) compared the accuracy of IMU sensors placed on the left-wrist, right-wrist, and the waist in a dining activity. In their results, they found that although the combined values from all three accelerometers had the highest accuracy (81%), readings from the dominant hand sensor had comparable accuracy (80%). Likewise on a per-activity basis, no single placement consistently outperformed the other placement. Bao and Intille (2004) compared the difference in recognition accuracy of sensors placed at the hip, wrist, arm, ankle, and thigh for a mixture of 20 low and high level activities[3], and found that only using sensors from 2 locations on the body was almost equally as effective as using all 5 locations. Given this information, there is no optimal set-up for sensor configuration. Additionally, Huang et al. (2017) and Bao and Intille (2004) have demonstrated that placing more sensors may not necessarily significantly improve recognition accuracy. Another component to consider for sensor numbers and placement is the comfort of the participant; in order for a sensor set-up to be feasible in realistic conditions, the participant should not need to be encumbered with sensors on a daily basis, and they should not be expected to wear sensors in the exact same position and orientation for each instance of recording (Andreu, Baruah, and Angelov, 2011).

Finally, sensor placement must also consider the possibility of displacement, meaning the ability of the sensor to change location and/or orientation from their intended position overtime. Many machine learning algorithms are written based on the assumption that the placement and orientation of the on-body sensor is fixed, but overtime small displacements via clothing friction or loosening of the grip mechanism (e.g Fastners coming loose, adhesive material dissipating) may induce displacement or disorientation (Kunze and Lukowicz, 2014). In instances where sensor placement is expected to change, orientation detection or placement independent algorithms need to be developed.

[3]Activities monitored: walking, sitting, standing, watching TV, running, stretching, scrubbing, folding laundry, brushing teeth, riding elevator, walking while carrying items, working on computer, eating/drinking, reading, cycling, strength-training, vacuuming, lying down, climbing stairs, riding an escalator

4.2.4 Feature Engineering

With feature engineering, the aim is to generate and select properties (called 'features') from raw signal data. This section covers what hand-crafted features are typically selected for HAR. Features can also be computed through a process called feature learning. However, as feature learning is an integral component of deep learning neural networks, this process will be described further in Section 4.2.6 instead.

4.2.4.1 Hand-Crafted Features

The process of selecting features manually (often referred to as the "hand-crafted" approach) entails that the features are chosen based on the researcher's experience of the domain. The features obtained in this method can be understood in concept by human interpretation. Because of the simplicity of the approach, the calculation of features is often computationally inexpensive; linear calculations like the mean, root mean square, and zero-crossing rate, can be calculated in a matter of 20µs with average processing speeds (Gao, Bourke, and Nelson, 2014). This makes the approach ideal for low-cost or real-time HAR systems where computational speed can be a decisive factor. However, despite being the simplest approach to feature extraction, the choice of selecting appropriate hand-crafted features for a HAR system can be very challenging task.

Although statistical components of the time domain are very frequently used, few investigations try to justify their inclusion. Espinilla et al. (2018) reflects this observation; when testing 27 time-frequency domain features across different activity monitoring scenarios, less than 50% (11/27) of features were found to be relevant across all scenarios after applying dimensionality reduction. One of the included features; total energy, was irrelevant in all scenarios. This demonstrates that a broad inclusion of features can actually be detrimental to the performance of the system, and further highlights the importance of applying dimensionality reduction methods. In contrast, Zheng (2015) took a more calculated approach to their selected features; to recognize their 6 measured activities (walking, jumping, running, standing, sitting sleeping), they state that because muscles produce different forces during different states of activity (Cola, Vecchio, and Avvenuti, 2014; Cross, 1999; Hof, Van Zandwijk, and Bobbert, 2002), these forces proportionately produce different accelerations via Newton's Second Law, this meant the mean magnitude of acceleration could be used as a feature. They further justified including the tilt of the accelerometer and entropy as additional features by reasoning that random walking could be represented by turning angles (Wu et al., 2000), and that Shannon entropy was a measurement of signal uncertainty of the accelerometer in the time domain. With a single accelerometer, they used these features to achieve an average of 95.6% in HAR accuracy. This investigation showed that, with the proper research and fundamental understanding of differences in their physical activity properties, high performances can be attained only using one sensor and a small but well-balanced feature set.

Data is often translated from the time domain via Fast Fourier Transform (FFT) to acquire an expanded range of features in the frequency domain. Chen et al. (2018) argues the importance of frequency domain features by reasoning those different activities will have different base frequencies which are distinguishable in the frequency domain. Bao and Intille (2004) concurs, explaining that frequency domain features like spectral entropy and dominant frequency analysis can help to differentiate activities that have similar levels of energy, with the example given being that cycling can be differentiated from running; cycling will have a dominant frequency magnitude in the vertical direction compared to running, while running will have higher spectral entropy. Despite having more features to work with, adding frequency-domain features may not necessarily result in improved performance. Rosati, Balestra, and Knaflitz (2018) tested two feature sets: feature set A contained only features from the time domain, while feature set B contained combined features from the time-frequency domain. They found that using feature set A resulted in marginally better overall classification performance (97.1 vs. 96.7%)[4], and when tested with certain classifiers, feature set B had a significant 40% loss in accuracy. Rosati, Balestra, and Knaflitz (2018) conceded however that feature set B gave better understanding of biomechanical behaviour particularly in participants with pathological gait. Tangentially related to the frequency-based features are Mel-Frequency Cepstral Coefficients (MFCCs), which were originally conceived for audio processing applications (Logan, 2000). In short, these features are the coefficients obtained from the cepstrum (a spectrum generated from the inverse of a Fourier Transform) of a Mel-frequency scale (a logarithmic frequency scale centred around the frequency region of interest). The benefit of acquiring these coefficients, in the context of HAR, is that, because there is an increased resolution around the lower end of the frequency scale, which is typically the base frequency of most human activities, the MFCCs can effectively capture more detailed information about the activity as opposed to a standard frequency scale. MFCCs have been applied in a handful of HAR papers using accelerometer data with resounding success (Batool, Jalal, and Kim, 2019; San-Segundo et al., 2016; San-Segundo et al., 2018; Takeuchi et al., 2009).

There are many other extractable features which do not necessarily pertain to time and frequency domains. Preece et al. (2009) for instance utilised features in the Wavelet domain. Signals in the wavelet domain are represented by wave-like oscillations ("wavelets") that begin with zero amplitude, temporarily oscillate, then reduce back to zero magnitude. The coefficients used to describe the behaviour of these wavelets (or more specifically, statistics relating to the wavelet coefficients) are occasionally used as features in some investigations Mantyjarvi, Himberg, and Seppanen (2001), Najafi et al. (2003), Sekine et al. (2000), Sekine et al. (2002), and Song and Wang (2005), however Preece et al. (2009) is one of few studies to compare using exclusively wavelet domain features with time/frequency domain features. The resulting wavelet feature set had lower sensitivity and specificity across all 8 tested

[4] Activities monitored: resting, standing, level walking, upstairs, downstairs, uphill, downhill

activities[5] compared to a feature set comprised of time and frequency features; however this only suggests that exclusively using wavelet-domain features will not build a robust classifier, and that wavelet features may still be viable as complementary components to the main feature set. Discrete domain features (where data is represented as symbolic string data) were briefly explored by Figo et al. (2010) using Dynamic Time Warp, Levenshtein Edit Distance (Gusfield, 1997), and the Euclidean distance between symbolic string representations (Lin et al., 2007). Unfortunately, these feature sets performed very poorly when recognizing between just 3 activities (walking, running, and jumping), with the Levenshtein Edit Distance having the highest accuracy of 52.5%, which would suggest discrete domain features are not well suited to HAR problems.

4.2.5 Dimensionality Reduction

Dimensionality reduction is when features are "compressed" such that redundant features are removed and/or combined with each other (depending on the approach) to form more informative features. As more and more features (the "dimensions") are added to a machine learning system, algorithms can begin to suffer from a phenomenon known as the "curse of dimensionality" (Bellman, 1966). Having too many irrelevant features can lead to overfitting and can impact adversely on the performance of a HAR classifier. By contrast, when the dimensionality of data is adequate enough to be represented in 2 or 3-dimensions, the feature data can be visualized, which allows for easier human interpretation of how different classes (i.e human activities) behave; in turn this can be useful to determine cluster composition in unsupervised learning approaches, which is discussed further in Section 4.2.6.2. Fig 4.1 demonstrates a simple example of dimensionality reduction. Two strongly correlated (and thus redundant) features, x_1 and x_2, can be projected into a lower dimensional space and represented by the transformed feature, z_1 with minimal loss of information.

[5] Activities monitored: walking, walking upstairs/downstairs, jogging, running, hopping (left and right leg), jumping

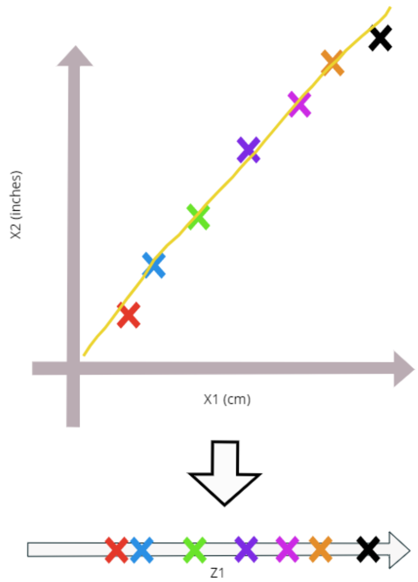


Figure 4.1 Example of reducing two dimensions with redundant information into a single dimension, recreated from Ng (n.d.).

In machine learning literature, *"feature selection"* and *"dimensionality reduction"* are often used interchangeably. For full clarity in this thesis, *"feature selection"* is considered to be a unique instance of *"dimensionality reduction"* in which the retained features are unaltered, whereas "feature transformation" involves changing (transforming) the values of the features to incorporate fewer dimensions whilst retaining as much valuable "information" from the original dimensions as possible.

4.2.5.1 Feature Selection

Feature selection in HAR can be distilled into 2 approaches: the Wrapper method and the Filter method. Fig. 4.2 presents a workflow diagram of the filter and wrapper methods. In the filter method, filter algorithms will measure intrinsic properties that evaluate the relationships between variables, and based on these relationships it will establish the optimal subset of features for the algorithm. Wrapper algorithms on the other hand will seek to add or remove features (depending on the design of the wrapper) based entirely on the resulting

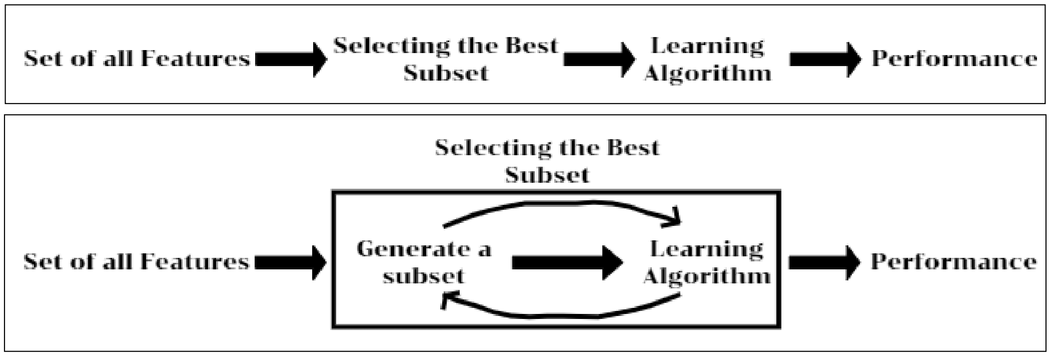


Figure 4.2 Top: flowchart of the filter feature selection approach. Bottom: flowchart of the wrapper feature selection approach. Recreated from Kaushik (2016).

classifier performance. Both approaches have their strengths and limitations: filter methods are not classifier dependent, while wrapper methods are, which can make the wrapper approach computationally expensive, especially if there are multiple classifier candidates being investigated. This has led to some wrapper algorithms being called "greedy" and with careless implementation can be considered a sub-optimal approach (Manini and Sabatini, 2010). This can especially be an important factor if the feature selection has real-time implementation (Karagiannaki, Panousopoulou, and Tsakalides, 2017). Conversely, as the wrapper approach gets feedback from the classifier algorithm, it is more likely to find the true optimal set of features than the filter approach.

Of particular note for this chapter are the Relief-F, chi-square and minimum redundancy Maximum Relevancy (mrMR) filter feature selection methods, and the Sequential Feature Selection (SFS) wrapper method. The Relief-F algorithm is a popular supervised filter-based feature selection algorithm; To rank features, all feature weights are initially set to 0. A random observation x_r is selected and updates feature weights based on the observation's proximity to the k-nearest neighbours of other observations x_q, the algorithm gives higher weights towards features of x_q that give different values to neighbours of different classes, while giving lower weight to features of x_q that give different values of the same class. This is repeated for all observations, and once completed, the algorithm ranks features by their highest combined weight across all observations.

The chi-square algorithm is another supervised feature selection algorithm, it determines whether a predictor variable and the output class are independent using statistical chi-square tests. Therefore, it uses a small p-value that rejects the null hypothesis to determine whether a response variable is independent on a predictor variable, with small p values indicating strong features. However, the algorithm is limited in its ranking score, which is calculated as $-log(p)$. Therefore, strong features that have small p-values can get output as infinite and if this happens to multiple features, it is impossible to compare their rank.

First described by Ding and Peng (2003), the mRMR feature selection process is a heuristic iterative process that calculates a score for each feature f and iteration i, through a relevance to redundancy ratio of the target variable (in this context, the activity class):

$$score_i\left(f\right) = \frac{relevance\ \ (f\ |\, target)}{redundancy\ \ (f\ |\, features\ selected\ until\ i-1)} \tag{4.1}$$

The method in which relevance and redundancy are calculated can vary depending on the software (Zhao, Anand, and Wang, 2019). Matlab's official documentation states that both variables are calculated based on the Mutual Information between the feature data and the associated class (Mathworks, n.d.[b]). As the process is heuristic, the top ranked features can change when the dimensionality and the size of the dataset are both large. This variance can be dealt with by applying cross-validation; wherein for each "training" partition, the mRMR algorithm is ran and scores are retained for each iteration. After all iterations are complete, the cumulative score for each feature across all iterations is calculated and the features with the 1st – nth highest score (as set by the user) are chosen as the relevant features. This also prevents the feature selection process from overfitting as it is not using the entire dataset to try and formulate the optimal feature set.

Wrapper filter feature selection methods like the SFS are mechanically simple but computationally expensive; a classifier function is applied to a feature set X that has a predictor variable y. Depending on the configured mode of operation, the SFS algorithm either begins with no features and sequential adds combinations of features that minimize the training loss as defined in the classifier function until a global minima is reached (forward SFS), or begins with the full set of features and sequential removes features until the global minima is reached (backward SFS). Regardless of the approach, the linear approach to optimization can lead to long training times when using a complex dimensionality dataset, and do not guarantee an optimal feature set will be reached regardless (Molina, Belanche, and Nebot, n.d.).

The filter and wrapper methods have equal viability in HAR, which is reflected in the literature. In Nguyen et al. (2018), two filter methods (Laplacian and Relief-F) were compared with a wrapper method (SFS) for activity recognition performance in multiple sensor positions on the body. In most sensor positions, the filter and wrapper methods both performed equally well, and in others the filter methods only marginally outperformed the wrapper by a 5% margin. Zhang and Sawchuk (2011) investigated how the number of features impacted on the feature selection algorithms (Relief-F filter, forward SFS and Single Feature Classification wrappers). While the Relief-F filter performed very poorly when only 5 features were selected, it had the best performance when 50 features were selected. Both filter and wrapper methods performed equally well in intermediate ranges of features. Suto, Oniga, and Sitar (2016)[1] and Suto, Oniga, and Pop Sitar (2016)[2] compared the same feature selection approaches (1 wrapper, 8 filters) on different datasets. In Suto, Oniga, and Sitar (2016)[1], the lone wrapper method had the best HAR performance, while in Suto, Oniga, and Pop Sitar (2016)[2], the chi-square filter performed better in more classifier algorithms. These examples demonstrate that, as with many other configurable options in machine learning, the choice of feature selection approach will be problem dependent.

4.2.5.2 Feature Transformation

4.2.5.2.1 Principal Component Analysis In feature transformation processes of dimensionality reduction, the aim is to reduce the number of dimensions (features) by combining the features into a lower dimensional space. The most well known and most widely used transformation process is Principal Component Analysis (PCA). The function of PCA is to represent the total variability of a feature dataset in a series of eigenvectors of a covariance matrix, known as principal components, with the ideal PCA representation using as few of these principal components as possible to capture almost all of the total systematic variance. As PCA may create data compression by limiting the number of components fed into a classifier (and thus reducing computational costs), it is a desirable technique in mobile computing (Wang et al., 2005). Chen et al. (2017) combined PCA with a co-ordinate transformation system which allowed a mobile phone to recognize activities regardless of its position on the human body. The processing time for PCA was 0.4ms per sample, making it suitable for real-time daily monitoring. PCA may not always benefit a HAR dataset. Walse, Dharaskar, and Thakare (2016) for instance took 561 raw features and via PCA reduced the number of components to 70. While the model building time was 500% faster using only PCA components, the accuracy of the system actually decreased from 98% with all features to 96% with PCA components. This could still be considered a very good performance but does indicate a trade-off when implementing the technique if the components that are not preserved still contain information relating to physical activity.

PCA is an unsupervised learning technique, which means that it does not require the “class” (i.e the activity) to perform its calculations. The supervised equivalent of PCA is called Local Discriminant Analysis (LDA). LDA performs dimensionality reduction in a similar manner to PCA, but the key difference between the two techniques is that LDA considers the class associated with each feature when obtaining linear relationships between components. This also means that LDA in of itself can be used as a classifier (Balakrishnama and Ganapathiraju, 1998). LDA has been proven to be superior to some standard feature selection methods (Yang et al., 2007), and was found to be the best performing base classifier in a study by Siirtola, Koskimäki, and Röning (2019) recognizing activities from a public dataset constructed by Shoaib et al. (2014).[6] However, Tao et al. (2016) demonstrates

[6] Activities monitored: walking, sitting, stand-ing, jogging, biking, walking upstairs and downstairs

that LDA and PCA had very similar performances in 2 different datasets, with LDA only marginally outperforming PCA by 0.003% overall. Further to this, as both LDA and PCA assume linearity between components, LDA has performed poorly in some investigations such as Khan et al. (2010), where non-linear discriminant analysis had a significant performance increase (60% for LDA vs. 96% Kernel Discriminant Analysis respectively).

4.2.5.2.2 Non-Linear Dimensionality Reduction Techniques In non-linear variations of feature transformation, it is no longer assumed that variables have linear relationships with each other. PCA has a non-linear variant which is commonly referred to as kernel PCA (kPCA). While linear PCA aims to transform data from high dimensional space to an orthogonal lower dimensional space, kPCA will initially represent the data in a higher-dimensional space which is defined by a kernel function. Fig. 4.3 demonstrates the effectiveness of kPCA when handling non-linear distributions of data.

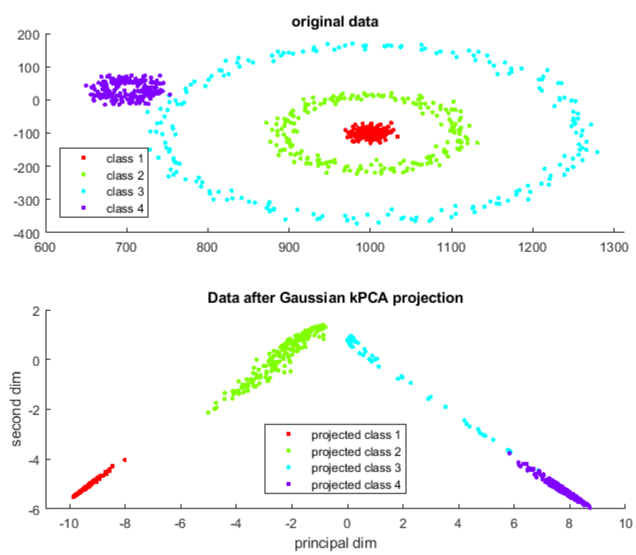


Figure 4.3 Example of kPCA projection with non-linear data. This image was generated in Matlab using code originally written by Kitayama (2021)

In order to translate into a lower dimension, the kPCA must first map the features in a higher dimension space. A kernel function mapped in the lower dimensional space is calculated such that it best fits a kernel matrix in the higher dimensional space, this process is known as the kernel trick. Then, to map onto lower dimensional space, the first K vectors are chosen to represent the transformed data, similar to choosing the number of principal components to conserve in linear PCA. A significant disadvantage of kPCA is that it scales poorly compared to PCA in large datasets; a kernel map requires $(n * n)$ dimensions where n is the number of observations. Ergo, having high numbers of observations can consume a considerable amount of RAM making it unviable when the memory of the processing device is limited. kPCA has shown instances of performing very well in some datasets while also underperforming in others. Hassan et al. (2018) demonstrated 89% mean accuracy in the best classifier across 6 activities[7] plus transitions between activities[8] using the kPCA process. kPCA also outperformed regular PCA in both El Moudden et al. (2018) and Wu et al. (2019). In the latter, a reduced form of kPCA was also proposed which effectively cut down on calculation time with minimal adverse impact on classification rate; thus, demonstrating kPCA can be used in mobile applications. kPCA performed poorly in Kästner, Strickert, and Villmann (2013), having similar projection output to regular PCA.

In addition to kPCA, there exist other non-linear dimensionality reduction techniques. t-distributed Stochastic Neighbour Embedding (tSNE) and Uniform Manifold Approximation and Rejection (UMAP) can bring greater separability of data clusters, making them a useful technique for unsupervised learning and have been shown to operate well in HAR contexts (Guo et al., 2016; Gupta, Mcclatchey, and Caleb-Solly, 2020). t-SNE was first conceptualized in Maaten and Hinton (2008), as an evolution of the Stochastic Neighbour Process (Hinton and Roweis, 2003). To give a simple overview, the t-SNE process iteratively reduces from the original number of dimensions of the feature matrix to the desired number of dimensions by first calculating the pairwise similarities of the points in the higher dimension. These pairwise distances become joint probabilities by applying a gaussian distribution; points

[7] Activities monitored: standing, sitting, lying down, walking, stairs

[8] Transitions monitored: stand-to-sit, sit-to-stand, sit-to-lie, lie-to-sit, stand-to-lie, lie-to-stand

that have high similarity in high dimensions will have higher joint probability. The points are then randomly assigned into a lower dimensional space. Similarities are once again calculated in the lower dimensional space but utilises a t-distribution to calculate the joint probabilities. The similarity matrices of the higher and lower dimensions are compared, and t-SNE undergoes a complex optimization process to try and match the similarity matrix of the lower dimension to the higher dimension process. The process then repeats until it reaches the desired dimensionality level. The t-SNE process can be superior at conserving complex relationships between variables than linear dimensionality reduction processes like PCA; however, its main drawbacks are the random initializations of the data points in lower dimensions, this causes the cost function in the similarity optimization process to be non-convex (there can be multiple optimal solutions), which can result in different t-SNE models for consecutive executions of the t-SNE process despite using the same data.

UMAP is a state-of-the-art dimensionality reduction technique, first introduced by (McInnes, Healy, and Melville, 2020). It operates with a similar fundamental objective to t-SNE, where it tries to map data in higher dimensions to a lower dimension by optimizing their similarities. The difference is that UMAP attempts to conserve the topology of higher dimensions by constructing lower dimensional elements known as “simplexes”. The points which are used to construct the simplexes are found through nearest-neighbour search radii, which will vary depending on the local density of the higher dimensional clustering region. Like t-SNE similarity calculations, points of a simplex object that are closer together in high dimensions are given higher weights for connection probabilities. The UMAP process then begins to map the dataset in lower dimensions by performing complex optimization processes to find the lower dimensional layout that can best conserve the topological structure of the higher layer. The mapping process exhibits “rotational symmetry”, which means that, like tSNE, the resultant low dimensional space can appear different when repeatedly initializing the algorithm on the same data.

A further drawback to both t-SNE and UMAP in their basic design iteration is that the mapping function is inherently non-parametric, meaning that unlike in PCA and kPCA, there is not a direct way to map testing data for classification. While parametric mapping capabilities for both approaches exist (Maaten, 2009; Sainburg, McInnes, and Timothy, 2020), their caveat is that the mapping functions require training a neural network to correctly operate. As such, these techniques are currently better suited for data visualization as opposed to being used for training classifiers.

4.2.6 Classifiers

This section covers the 3 main approaches of classification: supervised learners, unsupervised learners, and neural networks. There exists a wide variety of variants and hybrid combinations of classifiers (e.g semi-supervised learning, reinforcement learning, instance learning etc.) and it is beyond the scope of the thesis to cover all of them, the purpose of this section is simply to introduce the most common approaches of classification and their relevance to HAR.

4.2.6.1 Supervised Learning Classifiers in HAR

4.2.6.1.1 Support Vector Machines Support Vector Machines (SVM) are a popular choice for HAR classification. They are inherently binomial classifiers (can identify two classes) that can be made multinomial (can identify three or more classes) via the application of the one vs. all or one vs. one meta strategies. In one vs. all, each step in the classification process compares one class to all other classes simultaneously and selects the class i that maximises the value of the hypothesis function $h_{\emptyset}^{(i)}(x)$. In one vs. one classification, each classification step involves a binomial classification calculation between every binomial combination of each class. The class that "wins" the most across all steps is chosen as the selected output class.

Fig. 4.4 gives a visualization of the SVM process on a low-dimensional sample dataset.

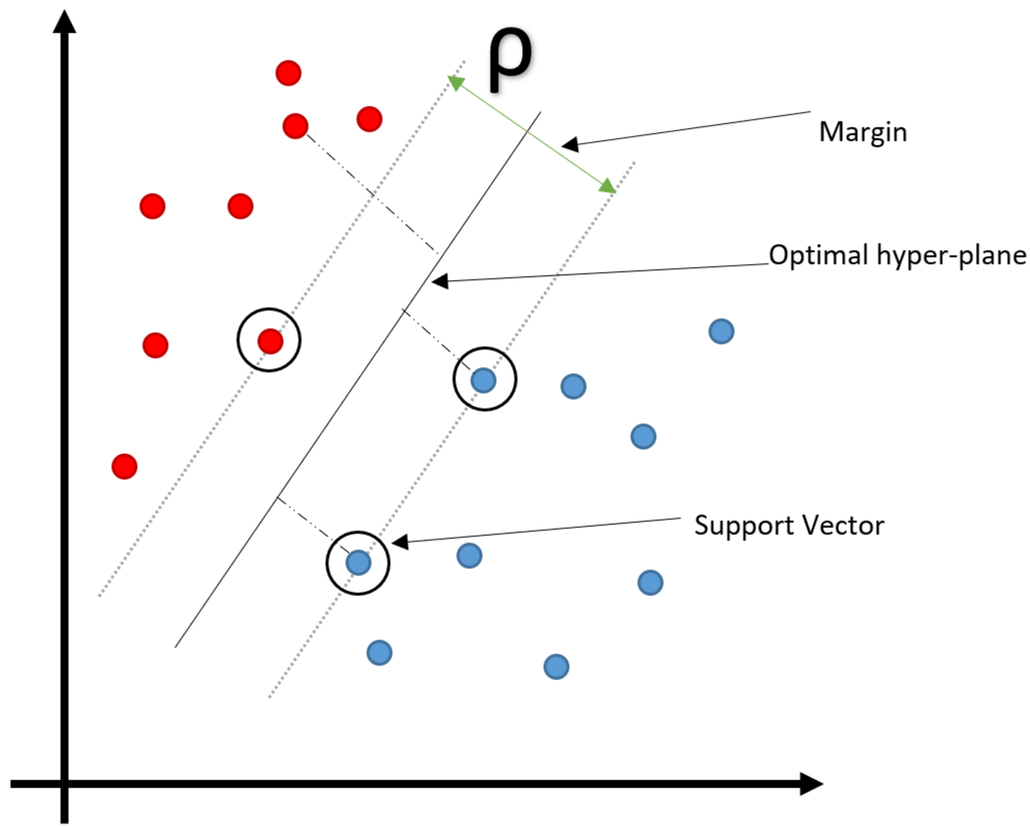


Figure 4.4 Visualisation of SVM classification, recreated from mitosis (2020)

The SVM algorithm chooses a small handful of input samples that are placed close to each other on a feature map and projects vectors from these samples through a process known as quadratic optimization, where the direction of the vectors is designed such that the perpendicular distance between the vectors, which are termed the "support vectors", have the maximised margin length ρ. The intermediate distance between the support vectors dictates the optimal hyper-plane, which is what ultimately provides the classification for new input samples. In these cases, a soft margin SVM is applied. The key difference between soft and hard margin SVM is that soft margin SVM introduces a slack variable which allows input samples to exist within the margin space. By compromising some input sample misclassifications, the margin of the SVM does not need to decrease to accommodate for all input samples and thus retains an intrinsic resilience to outliers. If the soft margin classification further fails to give acceptable decision boundaries in a linearly separable fashion, then the kernel trick should be applied. Typically, the Gaussian kernel is used as the mapping function, while linear, cubic, and polynomial kernels are also viable alternatives (Awasthi, 2020).

Due to its nonlinear surface mapping capabilities and soft margin component, the SVM is a flexible classifier and as such has seen frequent use in HAR datasets. SVM has also demonstrated powerful processing efficiency; Anguita et al. (2012) successfully employed an SVM on a mobile device for HAR, achieving 89% recognition accuracy on a mobile-friendly algorithm. Shdefat, Halimeh, and Kim (2018) compared a variety of SVM models (using different Kernels) with a variety of k-Nearest Neighbour (kNN) models for a smartphone-based dataset[9], and it was found that the SVM models on average performed slightly better than the kNN models, with Fine Gaussian SVM providing the highest accuracy out of all the models. Finally, Yin, Yang, and Pan (2008) applied SVM to detect "abnormal" activities, i.e activities that would not normally be expected during activities of daily living. These included slipping on the ground and falling down. To mitigate the lack of abnormal training data, they first trained the model on normal activities (walking / upstairs / downstairs and running) so that it could filter out potential abnormal activities. When the SVM was combined with a kernel non-linear regression model for abnormal activity filtering, it had a very strong Area Under Curve (AUC) score of 0.985, indicating a good balance between the sensitivity (true positive detection rate of abnormal activities) and specificity (true negative detection rate of normal activities). These articles have demonstrated the diversity and viability of the SVM algorithm.

4.2.6.1.2 K-Nearest Neighbour The k-Nearest Neighbour algorithm classifies an input by placing the sample on a feature map. In Fig. 4.5, a green circle representing a new unknown input is placed on a simple 2D feature map comprised of 2 classes, which are termed "red" and "blue" classes.

[9]Activities monitored: Attach to Table, Cycling, Laying, Pushups, Running, Sitting, Stairs, Standing, Walking, Jogging, Jumping

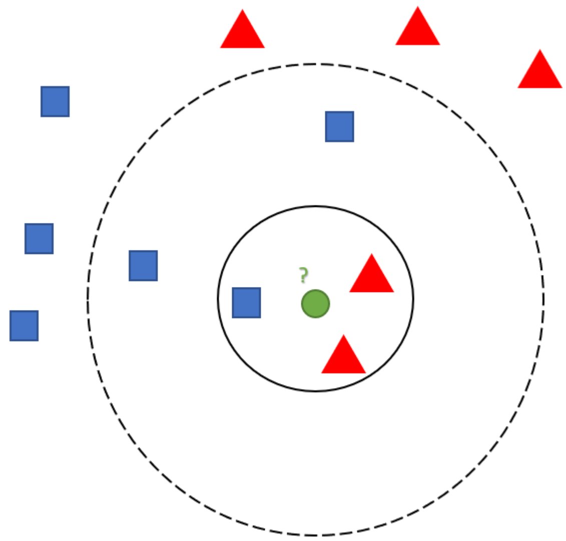


Figure 4.5 Illustration of kNN classification, recreated from Ajanki (2007)

The kNN algorithm calculates the distance (usually Euclidean, though other metrics are configurable) from the sample point to the other training features. The 'k' component of the kNN classifier dictates how many of the closest training samples (or "nearest neighbours") are evaluated in order to classify the new input. The choice of 'k' can be derived empirically or through analytical approaches such as the Elbow method (Yuan and Yang, 2019). Classifications are determined by the the class that has the largest representation in the trained nearest-neighbours.

In HAR literature, kNN has been used in a variety of studies. Like SVM, kNN has been used for smartphone-based activity monitoring (Kaghyan and Sarukhanyan, 2012). Further support of kNN in mobile computing was found in Ignatov and Strijov (2016), where a kNN classifier in combination with a novel segmentation technique had the lowest processing rate relative to the size of the dataset, outperforming other commonly used classifiers such as decision trees, SVM and logistic regression. Additionally, the kNN algorithm also had very high recognition performance for low-level activities[10], achieving an average of 96.6%. Notably, the authors remark that kNN is not a naturally fast algorithm and was only able

[10] Activities monitored: walking, jogging, stair climbing, sitting, standing

to achieve fast processing times through their segmentation process which generated short time windows of 1 second, indicating that kNN should be applied carefully in realtime HAR datasets. Sani et al. (2017), who in addition to using kNN as classification, also used a kNN sampling approach. The centroid of labelled data from a user is placed on a feature map containing all training instances. Instead of performing classification using all training instances, the algorithm then selects a subset of the training instances using the kNN algorithm (i.e it selects the K-nearest number of training samples), which are then subsequently classified using a variety of classifiers (kNN again, SVM or a neural network). The theory behind this approach is that since it takes a subset of samples based on a small number of labelled inputs, the training samples included for each participant are personalised to the user. This helped the classifier recognize different walking speeds (walk fast, walk normal, walk slow) as each training subset would be dependent of the user.

4.2.6.1.3 Decision Trees and Random Forest A Decision Tree (DT) breaks up the classification process into a series of comparative calculations (termed "decision nodes") that relate to the features of the input sample. The DT begins with a root decision node, and then depending on the outcome of the decision, splits into different branches with more decision nodes. The process repeats until the classifier reaches a leaf node, which indicates that it has enough information to provide a classification for the input sample. As the number of splits in the DT increases, the probability of the DT overfitting increases. Thus, a commonly applied technique is to utilise a Random Forest (RF). These classifiers are an extension of a DT classifier, consisting of randomly generated Decision Trees. Each DT will have randomly selected variables which are used to construct its behaviour. These decision trees will attempt to classify an input sample based on its tree construction, and once all results from the trees have been aggregated, the class that is chosen the most times is the decided class for the input sample. Fig. 4.6 visualizes the process.

As the decision trees of the RF utilise random variables, they are considerably less prone

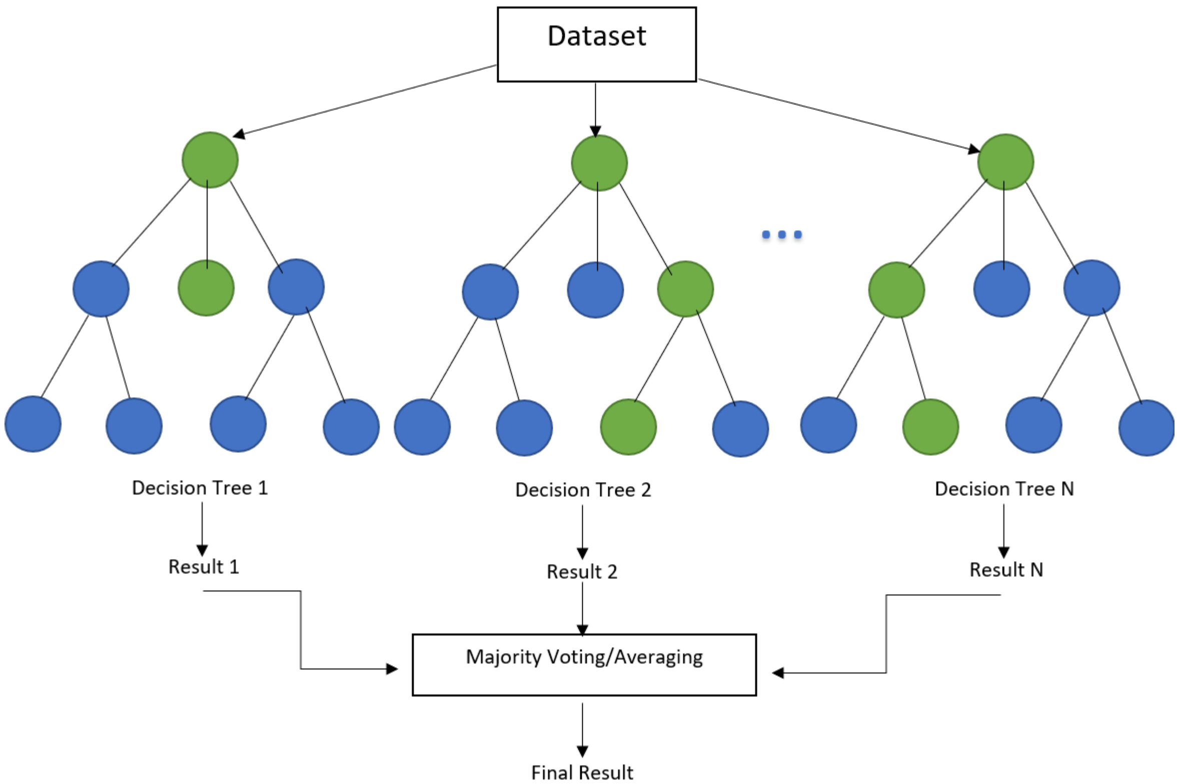


Figure 4.6 Illustration of RF classification, recreated from Sharma (2017)

to overfitting than standalone Decision Trees. However, because the process of variable selection is unknown to the user, a RF classifier can be considered a black-box machine, with the inner decision-making processes difficult to decipher.

RFs have been utilised numerous times in HAR. In Xu et al. (2017), RF classifiers outperformed singular DT and Neural Network classifiers in recognizing low-level activities[11]. They further noted that Random Forests were capable of cluster analysis and anomaly detection, making them useful in unsupervised settings. Likewise, Manda, Devi, and Row (2017) found RF classifiers had the lowest error rate of sport activity classification (various sequences of dumbbell movements) compared to kNNs and Decision Trees. Based upon the fundamental operations of the 2 classifiers, and the literature presented in this section, RF classifiers should at very least be considered as a viable alternative to Decision Trees, if not

[11] Activities monitored: walking, up/down stairs, jumping, running, standing

altogether replacing them as the classifier of choice provided that Computer Processing Unit (CPU) memory usage and training time are not of vital concern.

4.2.6.1.4 Ensemble Classifiers RF classifiers belong to a meta-classifier group called “Bagging Ensemble” classifiers. Ensemble classifiers describe any classification structure in which multiple classifiers are utilised. Generally, they consist of smaller classifiers that emulate the behaviour of the decision node on a DT; such that on an individual basis the small classifier might have high variance or high bias, which would lead to poor classification performance. This gives them their term “weak learners”. The ensemble classifier combines a weighted combination of the weak learners to form the “strong learner” ensemble classifier (Rocca, 2019). The majority of ensemble classifiers can be categorized into the “bagging” or “boosting” type. Bagging is a portmanteau of "Bootstrap Aggregation". The “bootstrapping” component of bagging signifies that a random selection of training samples is chosen with replacement (Efron, 1979). Meanwhile, the “aggregation” component of bagging signifies the output of each classifier is then evaluated and the overall classification is chosen. Boosting ensemble classifiers on the other hand train their weak models in a sequential fashion. After the first weak learner is constructed, the second weak learner builds on the first model and aims to reduce the classification error compared to the first one – the method by which it does this depends on the algorithm. The process repeats in a cycle until the boosting ensemble either achieves perfect training accuracy or the model reaches its pre-determined limit on the number of weak learners.

A popular example of a Boosting Ensemble classifier is the AdaBoost (AB) algorithm. With AB, the error of a weak learner is ameliorated by choosing a succeeding weak learner that minimizes the weighted sum error of the misclassification process. The overall weighted sum of the weak learners provides the final classification output of the model. In recent HAR literature, AB has demonstrated promising results. Subasi et al. (2018) investigated how commonly used classifiers performed individually on a public dataset when utilised as a weak learner in the AB ensemble[12]. Most of the included classifiers had some improvement in classification accuracy when implemented into AB, a notable combination was with a RF, which achieved a near perfect 99.98% accuracy. However, it is important to note that on an individual basis, the majority of classifiers already had very good performances – 7 out of the 9 classifiers had recognition accuracy of 98% or more. The only exceptional classifier was the Multivariate Alternating Decision Tree, which scored 84% individually and 98.8% when used with AB. These results could imply that the strength in the algorithms partially lay somewhere else - for instance, a strong feature selection – rather than strictly the implementation of AB.

4.2.6.1.5 Bayesian Networks and Naive-Bayes Classifiers Bayesian Belief Networks, also known simply as Bayesian Networks (BN) are a Probabilistic Graphical Modelling technique. That is to say, a BN can be represented by Directed Acyclic Graphs (DAGs). In the DAG of a BN, each node represents a variable (or series of variables) that are linked together through conditional probability – in other words the probability of an event B occurring given event A has already occurred. In practice, BNs can be difficult to formulate; they require the implementer to know or be able to derive the conditional relationships between their variables, and in turn this may require large volumes of training data to sufficiently derive a meaningful BN representation. A workaround is to use what is known as the Naïve-Bayes (NB) classifier. The main difference in a NB classifier is that it assumes that all variables are conditionally independent of each other. This assumption made by NB classifiers has resulted in inconsistent performance when implemented to solve HAR datasets. The activity set analysed in Lustrek, Cvetkovic, and Kozina (2012) was tested with NB, DT, RF and SVM. NB performed the worst out of the 4 classifiers regardless of the sensor's position, with classification accuracies ranging from 69.5 to 56.8%. By contrast, an NB clas-

[12] Activities monitored: working at computer, standing up, walking, stairs, standing, walking, walking and talking (to someone else), talking while standing

sifier used in Dengel et al. (2016) was the best performing classifier, outperforming SVM, DT and kNN. From the theoretical limitations of the NB classifier, the algorithm appears to be optimized when features are carefully chosen such that they have statistical independence. This can be seen in the work by Gupta and Dallas (2014) which implemented the Relief-F and SFS algorithms and compared performances when combined with NB and kNN. The SFS algorithm, which selected just 11 out of the 31 original features gave the NB classifier gave a much lower training set error and subsequently had very high mean classification rate (97.8%) for low level activities and posture transitions[13]. The implementation of NB and BN in HAR has therefore demonstrated that using NB can be unreliable without rigorous feature selection, and the implementation of BN by itself can be a very demanding challenge.

4.2.6.2 Unsupervised Learning Algorithms

The core difference between supervised and unsupervised learning is that unsupervised classifiers do not require labelled instances of data for training. Since the data is no longer labelled, an unsupervised classifier will instead classify an input sample as belonging to a 'cluster' of data. In the HAR domain, unsupervised learning is popular with vision and smart home based HAR, as the range of potential activities can exceed a preconceived list of prescribed activities (Ariza Colpas et al., 2020). However, there has also been plenty of research relating to HAR with clustering analysis. Some popular types of clustering in this field pertaining to HAR datasets include K-Means clustering, Hierarchical Clustering, Density-Based Spatial Clustering of Applications with Noise (DBSCAN) and Gaussian Mixture Model (GMM)-based Distribution Clustering.

4.2.6.2.1 Centroid Clustering and the K-Means Clustering algorithm In centroid clustering, data points are grouped according to the most mathematically representative data point for that cluster, which is known as the data centroid. Fig. 4.7 shows an

[13] Activities monitored: jumping, running, walking, sitting, sit-to-stand, stand-to-sit

example of organizing data into 2 clusters based on centroids.

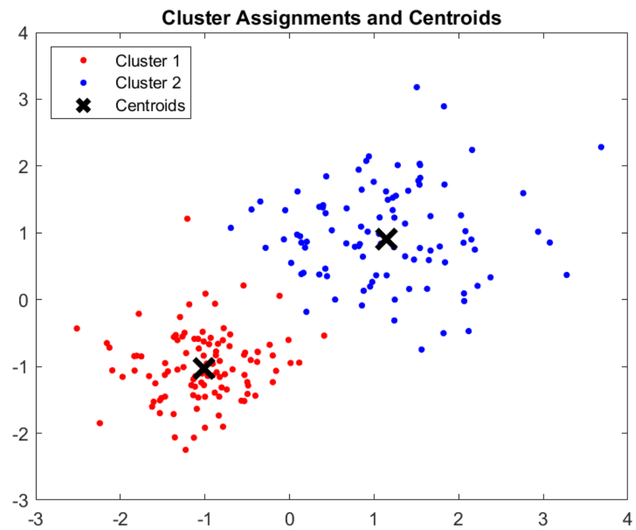


Figure 4.7 Illustrated example of centroid clustering

K-Means clustering is one of the most widely used clustering algorithms. It begins by initializing the data centroids in the feature space. The number of centroids, K, are determined by the user, hence the name 'K-Means Clustering'. There are several methods of initialization, but the most popular approach is the Forgy method (Peña, Lozano, and Larrañaga, 1999), which selects K training examples at random and uses their means as the initialization point in the feature space. The algorithm then undergoes an iterative procedure which loops between a clustering step and a centroid displacement step. In the clustering step, the algorithm assigns each data point in the feature space to the centroid with a distance metric. In the centroid displacement step, the centroids of the resulting clusters from the previous step are calculated, and the old centroids moves to these locations in the feature space. The procedure repeats these two steps until the point of convergence when there is no longer any displacement of the centroids and the within-cluster variance has been minimized, at which point the K-Means clustering algorithm is trained. In the testing component, a new sample point is distributed into the same space and is assigned to the nearest cluster by calculating the distance from the point to each cluster, selecting the cluster with the smallest distance. Choosing the number of clusters in K-Means clustering is comparable to selecting the K

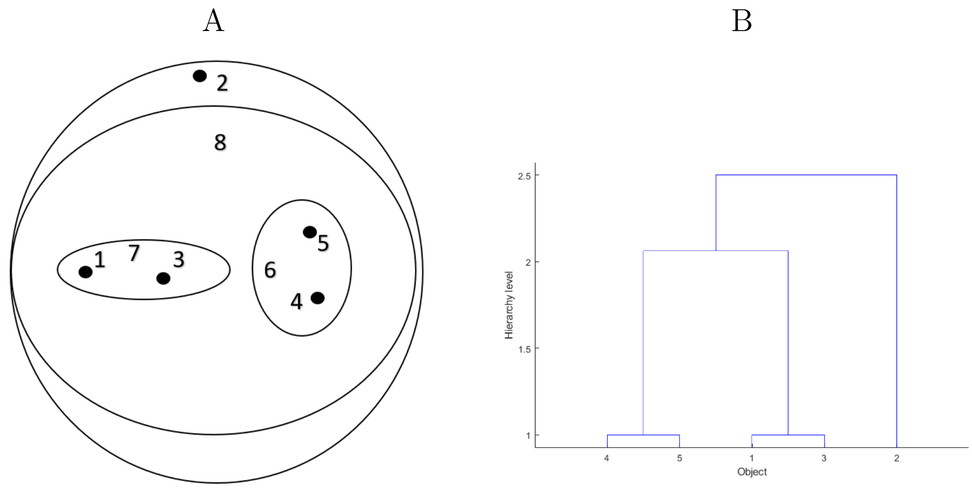


Figure 4.8 Visualization of hierarchical clustering. (A) represents the 'closeness' of data points. (B) is a dendrogram that represents how these data points would be organized in hierarchical conglomerations. Recreated from Mathworks (n.d.[c]).

in k-NN algorithms; meaning that the number of clusters can be chosen empirically, or by using methods such as the previously described Elbow method, or alternatives such as the average silhouette method and gap statistic method (Rousseeuw, 1987; Tibshirani, Walther, and Hastie, 2001; Yuan and Yang, 2019).

4.2.6.2.2 Hierarchical Clustering There are two main approaches to hierarchical clustering: Agglomerative and Divisive. In agglomerative clustering, each data point is initially treated as its own cluster. This starts an iterative process: in each step, proximity matrices are calculated to determine which data points in the feature space are closest to each other. Clusters with the highest proximity to each other are merged into a larger matrix. The process repeats until all points in the feature spaces have been merged into a single cluster. The Divisive approach is the opposite of the agglomerative approach, wherein a single cluster is broken down into individual clusters. Hierarchical clustering is interpreted through dendrograms, which show the iterative merging procedure that occurs with each cluster. Fig. 4.8 shows a simple example of how a dendrogram would be formed from hierarchical clustering.

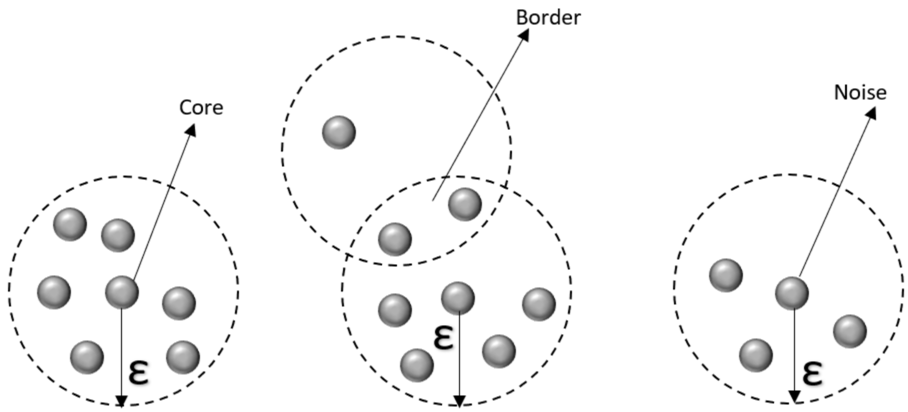


Figure 4.9 Illustration of DBSCAN algorithm identifying core, border and noise points. Recreated from Amini, Wah, and Saboohi (2014).

4.2.6.2.3 Density Clustering and DBSCAN Clusters in density-based clustering are created by dense formations of points in the feature space. This type of clustering is considered synonymous with the DBSCAN algorithm. It initializes by selecting a random data point. A density cluster is either formed or not formed depending on two controllable parameters: the first criterion, ϵ, is the minimum distance between points in the feature space, and the second criterion, $MinPts$, controls how many data points need to be in proximity to each other to form a cluster. If a data point has surrounding data points that are greater in magnitude than $MinPts$ and have a distance less than ϵ, the data point becomes a "Core point" and forms a cluster. If the data point fails to have neighbouring data points more than $MinPts$ and at a distance greater than ϵ, the point becomes a "Noise point" and is treated as an outlier. Once a core cluster has been formed, the DBSCAN algorithm iteratively expands its cluster by continuing to look for more neighbouring data points. As the cluster increases, it may encounter data points that are within the ϵ distance but do not have neighbouring points greater than $MinPts$, these points are called the "Border points". The Border points can be used to create better spatial definitions of the clusters (Amini, Wah, and Saboohi, 2014). Fig. 4.9 gives a demonstration of the identification process in DBSCAN.

The algorithm will iteratively expand the clusters until all data points have been classified

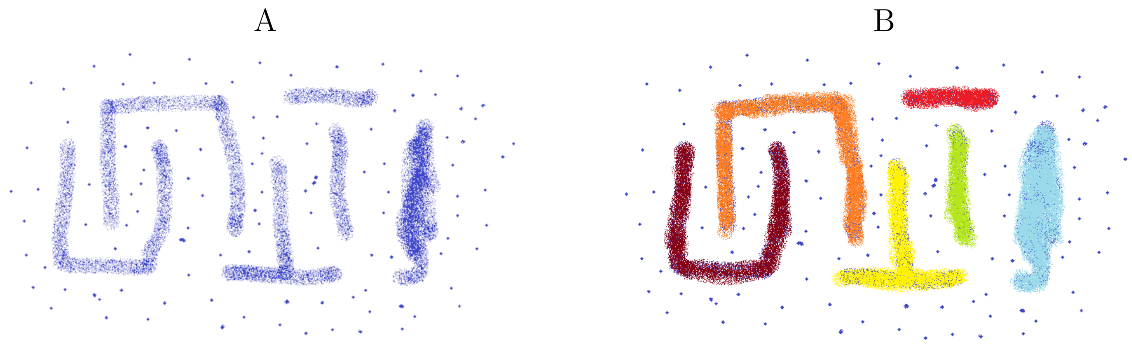


Figure 4.10 Visualization of DBSCAN clustering. A: Data points pre-clustering, B: Data points post-DBSCAN clustering. Recreated from Ivan (2018).

as a Cluster, Border or Noise datapoint. In Fig. 4.10, it can be seen that DBSCAN can be effective at handling non-linear distributions of data and discarding outlying data points. On the left-hand side (A) is the non-clustered data distribution, and the right-hand side (B) shows the data clustered under DBSCAN, with each unique cluster represented by a different colour and the dark blue data points representing noise samples.

4.2.6.2.4 Distribution Clustering and Gaussian Mixture Models The clustering approaches described so far have all incorporated some sort of distance or proximity metric to determine whether a particular data point belongs to a data cluster. Distribution clustering algorithms on the other hand incorporate a data point into a cluster based on its probability of being in that cluster, as such the data points are not necessarily "hard-assigned" to a cluster. The most well-known and widely used distribution clustering algorithm is the Gaussian Mixture Model. GMMs are a type of mixture model, which means that the feature space can be represented by multiple probability distribution functions. The clusters in GMMs follow a Gaussian distribution behaviour from the centroid. A typical plot of a GMM in a feature space will show probability 'bands', which indicate the confidence level of a data point belong to that particular cluster. Fig. 4.11 shows a simple example of a feature space under GMM distribution clustering. Three clusters are formed from 3 centroids, the darker colours indicate a stronger confidence value that the data point within that space will belong to that cluster.

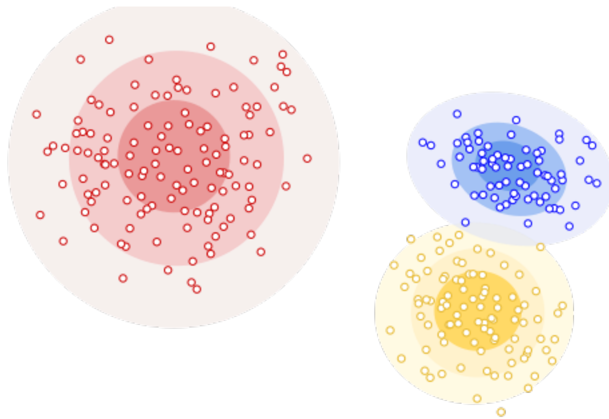


Figure 4.11 Illustration of GMM clustering from Google-Developers (2020). Image has C.C 4.0 license allowing free re-use.

The behaviour of the GMM (in terms of its centroid placement and distribution bands) can be derived through various methods, but the most commonly used is the Expectation-Maximization (EM) algorithm. The EM algorithm has an initialisation step: for a chosen value of K, the algorithm assigns random centroids μ_k with random variance σ_k^2 and a random value for the mixture coefficients, π (Keng, 2016). To clarify, the value of π signifies the proportions of the K neighbourhoods prior to beginning the EM algorithm. As with all other clustering algorithms described thus far, the EM algorithm undergoes a 2-step iterative process. The first step, "expectation", calculates the probability of each data point belonging to one of the μ_k clusters based on the distribution of the GMMs from the initial (or previous) step. The second step, "maximization", subsequently redistributes the GMMs such that it maximises its log-likelihood (i.e the GMM is redrawn to best fit the datapoints based on their probability of belonging to that cluster). The process repeats to the point of convergence when the calculated log-likelihood no longer changes between iterations.

4.2.6.2.5 Unsupervised Classifiers and Human Activity Recognition Unsupervised classifiers have found varying degrees of success in HAR investigations. K-Means clustering in Machado et al. (2015) had near-perfect cluster and classification accuracies of

99% each to recognize low level activities collected by a tri-axial accelerometer. Dinis (2018) constructed a hierarchical clustering model to detect a wide range of low-level and high-level activities[14]. Although the classifier was outperformed on average by an ensemble Decision Tree/Naïve Bayes classifier, the hierarchical method had better recognition performance on some of the included activities. Xia et al. (2020) was able to combine K-Means with DBSCAN to form a hybrid clustering model capable of recognizing specific motions in table tennis and badminton sports[15], attaining a prediction accuracy of 86.3%. Gani et al. (2019) used a GMM for low level activities[16], and acquired a perfect 100% recognition accuracy using only the y-axis acceleration from an accelerometer, simultaneously outperforming the included supervised models. When trialled on a public dataset (Anguita et al., 2013), the accuracy fell to 90%, but nonetheless demonstrated the capability of clustering models in the HAR field. Kwon, Kang, and Bae (2014) evaluated all 4 of the described clustering methods on their self-created dataset using mobile smartphone accelerometer data. For the parametric clustering methods (K-Means, Hierarchical and DBSCAN), GMM had perfect accuracy for all activities[17] for low values of k. However when k was increased to a maximum value of 50, hierarchical clustering had a more resilient accuracy drop-off compared to the other 2 methods, indicating potential to be able to correctly cluster when more activities are considered. DBSCAN, which is non-parametric and cannot pre-determine the number of clusters, was able to achieve greater than 90% clustering accuracy by optimizing its MinPts and ϵ values. These results show that, even if a clustering model can obtain high accuracy, it may not necessarily generalize well if subjected to different conditions.

[14] Activities monitored: sitting, standing, lying, walking, walking ascending stairs, descending stairs, running, cycling, nordic walking, watching TV, computer work, car driving, vacuum cleaning, ironing, folding laundry, house cleaning, playing soccer, rope jumping

[15] Activities monitored: Walking, Table Tennis Service, Table Tennis Stroke, Table Tennis Spin, Table Tennis Picking Up, Badminton Service, Badminton Drive, Badminton Smash, Badminton Picking Up

[16] Activities monitored: Walking, Stairs, Running, Standing, Sitting, Laying, Elevator Use and Driving

[17] Activities monitored: Walking, Running, Sitting, Standing and Lying Down

One of the advantages of unsupervised learning is that it can potentially extract more useful information than with supervised learning. As the data is no longer confined to a defined activity label, a robust clustering algorithm may identify unique characteristics which could belong to a previously unconsidered physical activity. They can also be further combined with supervised learning classifiers; Garcia-Ceja et al. (2014) for instance used a hierarchical clustering model to sub-class activities to counter intra-class fragmentation. To elaborate, certain activities could have very different behaviours when carried out in different contexts. Giving a simple example, say a researcher was interested in recognition of a person using a computer. While most people use a computer sitting down, some people may have access to standing computers which could change the acceleration behaviour of a worn sensor – this would require two different sub-classes which would be difficult to account for when constructing the study. When implemented, the hierarchical clustering technique should theoretically capture these sub-classes by grouping them together in the hierarchical model. This research demonstrates there is a lot of viability in employing unsupervised learning, even if the overall system still sometimes requires supervised classification.

The unsupervised approach comes with an inherent drawback in that evaluation of the clusters is more abstract than in supervised learning, where evaluation is linear. There are two forms of evaluation in unsupervised learning; extrinsic and intrinsic evaluation. Extrinsic evaluation of unsupervised clusters involves knowing the true label of the activity and to evaluate clusters of data, similar to supervised learning evaluation. Such techniques include Normalized Mutual Information (NMI) and the Adjusted Rand Index (ARI). This is directly informative evaluation of the clustering quality, especially when a clear clustering objective has been set (i.e clusters should correspond to activities), however unsupervised learning is typically deployed when the cost of annotation is too high to carry out, so will not always be a usable metric. Intrinsic evaluation methods on the other hand are always applicable to any unsupervised learning problem, as they evaluate the quality of the clusters themselves. Such metrics include the Silhouette score and the Dunn index.

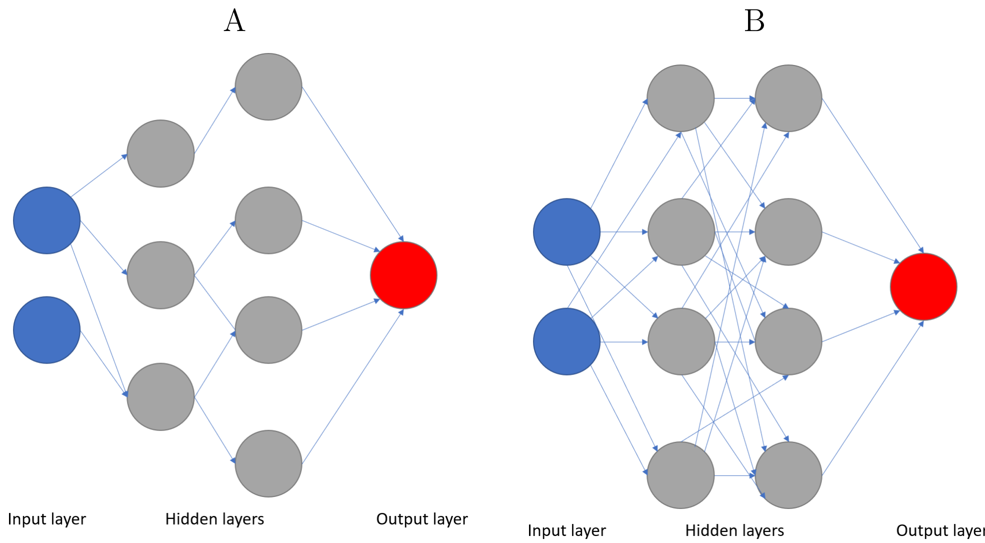


Figure 4.12 A: Topology of Feedforward Network. B: Topology of a Multilayer Perceptron. Recreated from Alex (2020).

4.2.6.3 Artificial Neural Networks and Deep Learning

4.2.6.3.1 Feedforward Neural Networks and Multi-layer Perceptrons Neural networks, formally known as Artificial Neural Networks (ANN) when applied in a machine learning context, are a powerful tool for classification. Feedforward neural networks (FF) are the simplest implementation of a neural network. In FF networks, information is only passed in one direction, from the input nodes to the hidden layer(s) and finally to the output nodes. As such, they can be represented graphically through DAGs much like Bayesian Networks.

A single layer perceptron connects all the input nodes in a neural network layer to the output nodes. A Multi-Layer Perceptron (MLP) is both an extension of a single layer perceptron and a special type of feedforward neural network, wherein every single node connects to each node in the next subsequent layer. MLPs are further defined by having the same hidden layer size for all hidden layers and have the same activation function. As they are an extension of an FF network, they can also be represented through DAGs. Fig. 4.12 demonstrates the structural difference between FF and MLP networks:

In literature, MLPs were generally found to have more representation in HAR literature compared to their FF counterparts. They are capable of recognizing low-level activities with high accuracy; the models constructed in Khan, Lee, and Kim (2008) had 99% accuracy for lying, standing, walking, and running activities. Bayat, Pomplun, and Tran (2014) found MLPs to be the strongest individual classifier compared to SVM, Decision Trees and Logistic regression classifiers for recognizing a variety of dynamic physical activities. An instance of MLPs performing poorly was seen in the work by Joshua and Varghese (2011), although it performed the best out of the three tested classifiers (the others being NB and DT), the MLP was still only able to achieve at best 80% accuracy for just 3 high-level builder activities (fetch and spread mortar, fetch and lay brick, filling joints), these results seem to suggest that the MLP works best when used for low level activity recognition.

4.2.6.3.2 Convolutional Neural Networks Convolutional Neural Networks (CNN) are an evolution of MLPs. Fig. 4.13 demonstrates a typical CNN architecture from Lee, Yoon, and Cho (2017) for classifying human activity based on 1 dimensional data from body worn sensors.

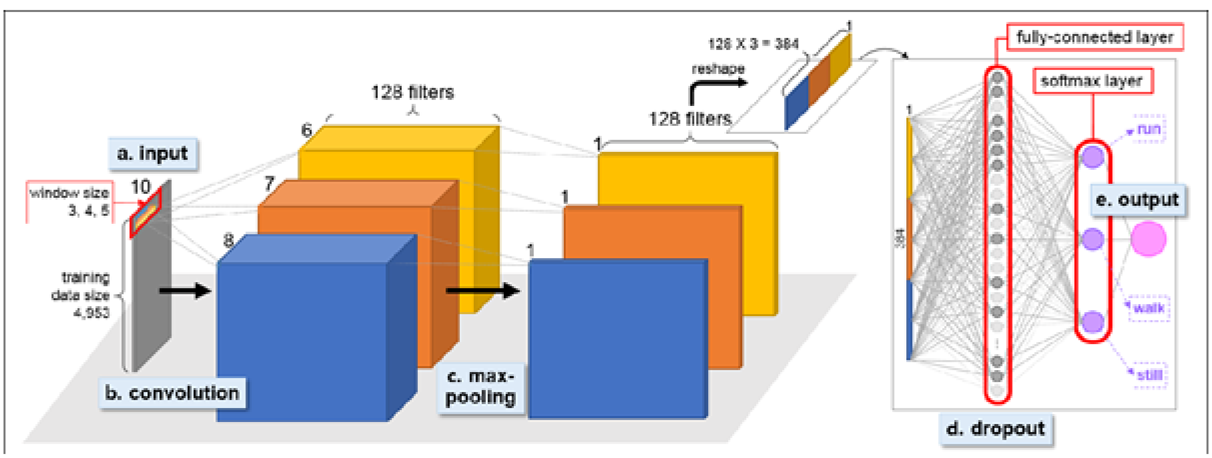


Figure 4.13 Overview of a CNN design, adapted from Lee, Yoon, and Cho (2017), © 2017 IEEE

CNNs contain at minimum 3 types of core layers: the convolutional layer, the pooling layer and the fully-connected layer. In the convolutional layer, the input signal is mathematically transformed via a convolutional kernel, with the transformation process varying

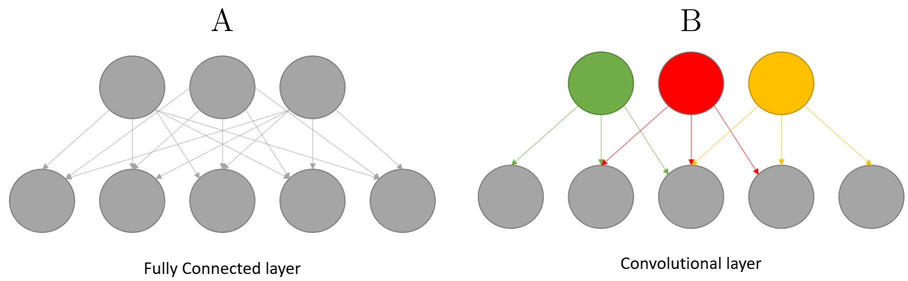


Figure 4.14 A: Topology of a Fully Connected layer. B: Topology of a Convolutional layer. Recreated from Despois (2018).

depending on the type of kernel used. Convolutional layers differ from the fully connected layers seen in MLPs, as they utilize shared weights between neurons to cut down on the number of parameters needed to construct the network, as illustrated in Fig. 4.14. This makes them more computationally efficient and better at recognizing patterns than with a MLP.

In Lee, Yoon, and Cho (2017)'s CNN, 128 kernels with varying window size of 3-5 gets passed through the raw data, which is segmented into 10 second chunks, this process creates 3 new feature map vectors which are 6-8 units in length. This process is what is referred to as "feature learning" and is what allows deep neural networks to perform recognition without the requirement of hand-crafted features. In the pooling layer, the input feature data has its spatial size reduced, which is done in order to cut down on the number of parameters and hence reduce the overall computational time. A common pooling technique as shown by Lee, Yoon, and Cho (2017) is the max-pooling process. The maximum value in each quadrant of the feature vector is chosen as the representative value for that quadrant, thereby cutting down all feature vectors in the example down to a length of 1. The resulting pooled vectors may then undergo several more iterations of convolution and pooling which can be used to generate further features for the classifier. Once the convolution and pooling processes have been completed, they are passed to the Fully Connected layer. Because of the previous convolution, this Fully Connected layer requires less parameters than would be required in an equivalent MLP design. Finally, most CNNs are classified in the output function typically via a softmax classifier. The softmax classifier uses the softmax activation function to calculate the cross-entropy, which determines the probability of the output belonging to each class, and the class with the highest probability is chosen as the predicted output. CNNs are traditionally designed for 2D image recognition processes, but their fundamentals are also applicable to 1-3D accelerometer and other body worn sensor data time signals for human activity recognition. CNNs are almost always referred to as a deep learning neural network in modern literature as its structure inherently requires at least 3 hidden layers, and these layers are what engineers the features for the classifier – thus the CNN only requires raw data as the input signal.

In HAR datasets, CNNs have shown consistently effective implementation. From the previously referenced paper, Lee, Yoon, and Cho (2017), they were able to discern between running, walking, and standing still with 92.7% accuracy. This demonstrated CNNs working at a very basic conceptual level for HAR. Panwar et al. (2017) also used a dataset which involved recognizing arm movements from various tasks and achieved a near perfect 99.8% recognition accuracy. Jiang and Yin (2015) took a unique approach to the input configuration. While the previously discussed studies fed the data in as a series of 1D raw accelerometer or other body worn sensor signals, Jiang and Yin (2015) transformed the raw data into 2D activity images via a DFT transformation of the raw data. After the transformation, the 2D image is fed into the CNN like a standard image recognition process to build and test the model. The approach proved a success, scoring 95.18, 97.01 and 99.93% accuracy on the UCI-HAR, USC and SHO datasets. As features are naturally learned in the CNN architecture, this research has highlighted the importance of intelligent pre-processing techniques to build a robust CNN classifier.

4.2.6.3.3 Recurrent Neural Networks and the Long-Short Term Memory Network Recurrent Neural Networks (RNN) are among the most prominent deep learning

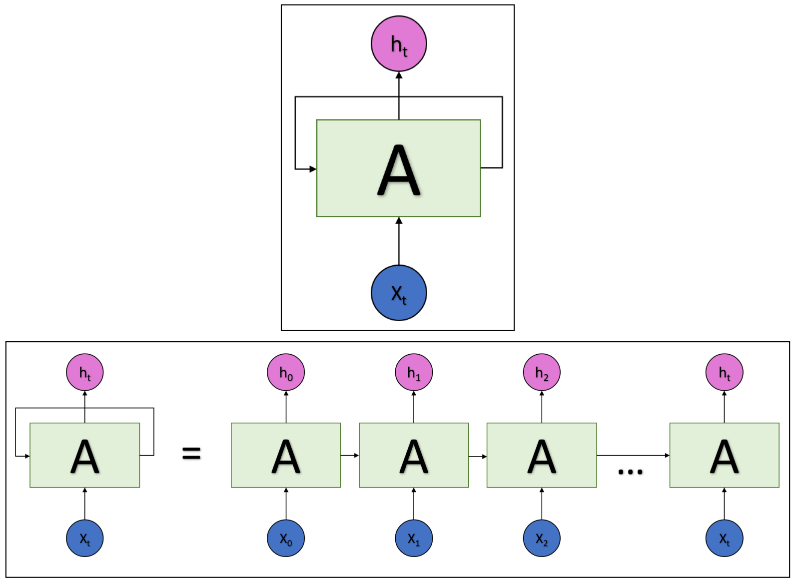


Figure 4.15 Top: Illustration of a rolled RNN unit. Bottom: Illustration of an unrolled RNN. Recreated from Olah (2015).

network designs. A RNN allows their hidden states to loop (or to "recur"). Fig. 4.15 (Top) illustrates the recurrence of an activation unit 'A' in the hidden layer. The output of the hidden state is looped back into the input of the hidden state, and so in the next iteration, the hidden state calculates its activation function based on inputs from the previous and current iteration, and the process repeats for the pre-determined number of loops. Fig. 4.15 (Bottom) shows how each recurrent loop can be symbolically unrolled into a linear sequence of activation units for 0 ... t iterations. At the end of the recurrence loop, the activation unit considers both the input x_t and all previous outputs $h_0 \ \dots \ h_{(t-1)}$ to generate its final output h_t. As a result of this looping, RNNs have a property called sequential memory which can be used to predict the "intent" of a sequential time-series input. The RNN is thusly capable of classification through the fast sequential processing of the sequence data, and as such does not require a formal feature extraction process as seen in supervised classifiers and other neural networks (Chen et al., 2016; Inoue, Inoue, and Nishida, 2018). This is again another example of "feature learning" as also seen in CNN architecture.

A disadvantage in standard RNN architecture is that when sequences in the loop are very long, the RNN has great difficulty making inferences of dependencies that are temporally far apart in the sequence, in other words it means the RNN is poor at recognizing long-term dependencies. This is known as the 'vanishing gradient' problem in RNN and is what led to them falling out of favour for machine learning solutions (Bengio, Simard, and Frasconi, 1994). In most modern RNN architecture, a variant called the 'Long-Short Term Memory' (LSTM) is used instead. LSTMs are structurally very similar to RNNs, with main difference being that in the recurrence hidden layers, they contain a memory cell known as the LSTM unit (Hochreiter and Schmidhuber, 1997). This LSTM unit has control over which components from the previous state are kept (and if so, to what degree) and which components are not included. This gives the classifier a much greater probability of recognizing long-term dependencies and thus build a more robust classifier. Once the recurrence loops are completed, the resulting loops are fed into the fully connected layer and passed to a SoftMax activation function to calculate the cross-entropy. This determines the probability of the output belonging to each class, and the class with the highest probability is chosen as the predicted output. Another frequently featured layer in LSTM and other Deep Learning networks is the Dropout layer. This layer controls the percentage of neurons from a preceding layer that are to be removed or "dropped out" from the training process. This creates a regularization factor for the network and prevents it from overfitting on data during the training process.

LSTMs have shown some promising results for HAR datasets. Chen et al. (2016) found an LSTM-based RNN network had 92% accuracy on the WISDM dataset. LSTM methods further outperformed Decision Trees, RF and SVM in Inoue, Inoue, and Nishida (2018) for the HASC and UCI-HAR datasets. Some research has presented less favourable results for the RNN approach; Zebin et al. (2018) compared multiple variations of the LSTM model on a self-collected dataset with six low-level activities[18]. Although the included models were all able to achieve high recognition accuracies (minimum 88%), all included models were

[18] Activities monitored: level walking, up/down stairs, sitting, standing, lying

outperformed by an SVM with standard feature extraction. They note in their discussion that this was likely because the dataset did not necessarily contain long-term dependencies which would have otherwise adversely impacted on the SVM. Milenkoski et al. (2018) found the LSTM being outperformed by a MLP model on self-collected datasets using both lab and field conditions and a mobile phone for accelerometer signals. While having very high prediction accuracy on walking, jogging, sitting, and standing, the LSTM visibly struggled at recognizing upstairs and downstairs activities (mean 97.2% recognition rate of jogging, sitting and standing vs. 58.7% mean recognition rate of upstairs and downstairs), it further had notably weak performance on a field collected dataset, with downstairs activities averaging at just 21% accuracy. From these datasets, it seems clear that LSTM performs best when using large datasets, and additionally works better when there are long-term dependencies in movement that the LSTM can recognize.

4.3 Part II: Interpretation of the review to construct a HAR study for ILLAs

The content of this chapter thus far has described the key design steps for a HAR study, and the various choices that can be made for the machine learning classification. The second part of this chapter is an extended conclusion that summarizes the key findings from the literature discussed, along with discussion of additional HAR studies conducted on populations with gait impairments (including lower limb amputations), to construct a novel HAR study.

4.3.1 The Data Source

At the beginning of the chapter, it was explained that public datasets can be a fast and cost-efficient method of acquiring data for a machine learning study. Unfortunately, regardless of how many activities are collected, the number of sensors used, or the number of participants recruited, publicly available datasets are not useful to the researcher if they do not contain relevant data. In this thesis, ILLAs are required for study. None of the datasets included in Demrozi et al. (2020)'s survey specifies inclusion of ILLAs and to the author's knowledge, no publicly available dataset exists which involves ILLA recruitments outside of a constrained laboratory setting. Hood et al. (2020) has a publicly available file on 18 ILLAs walking on a treadmill at different speeds, however this dataset is intended for gait analysis rather than activity recognition. Hu et al. (2018) collected IMU-based data in a laboratory environment using assistive robotics to emulate the movement patterns of an ILLA. Naturally, this is not as realistic as collecting data from amputees themselves, and the laboratory conditions further hamper the real-life applicability of the data. Thus, the study in this thesis acquired data first-hand.

4.3.2 The Data Collection Process

It was established in Section 4.2.1.2 that data collection for HAR can be controlled (in laboratory environments), naturalistic (free-living conditions), or a hybrid combination of both. The naturalistic approach was chosen for the activity monitoring study out of necessity: at the time of planning, the COVID-19 epidemic was present in the UK, and as a result many laboratories at the author's University (the University of Strathclyde) were closed throughout most of 2020. By conducting research in naturalistic settings, not only would the data collected be more representative of daily amputee movement, but it also meant it was possible to continue the investigations without major delays. Further to this, there are very few studies which actually conduct HAR with an amputee population in free-living environments. Arch (Schrank) et al. (2018) conducted a free-living experiment with ILLAs using a Fitbit® and StepWatch™, however their objective was to assess the accuracy of step counts, and not to try and recognize new activities. Gardner et al. (2016) developed an algorithm to differentiate sitting, standing and the donning/doffing of a prosthesis in free living conditions – this has limited scope in terms of capturable activities, and did not utilise machine learning. The strongest evidence showing a lack of relevant studies came from Labarrière et al. (2020)'s systematic review of HAR in amputee populations. They highlight that of 58 included studies, only two were conducted outside of laboratory environments (Zhang et al., 2019a; Zhang et al., 2019b). These studies both utilized visual information from a mountable RGB camera as a key component in their terrain recognition process, and as was emphasized at the beginning of the chapter (Section 4.1, including visual sensors opens up issues of privacy invasion as well as poor cost scaling. These findings indicated that this research would be one of the first studies to try and capture a wide range of activities from ILLAs using only wearable inertial sensors.

In addition to an ILLA population, it was decided to include individuals with no gait impairments in the study. This was done for several reasons: it was anticipated that, considering the low number of participants in the interviews of chapter 3 and the restrictions with recruitment due to the coronavirus epidemic, the number of amputee participants that could be recruited was expected to be small. Additionally, combinations of non-amputated and ILLA populations in HAR studies is not unprecedented, as seen in the literature review by Labarrière et al. (2020). By having a reasonably sized control population (estimate 6-10 people based on numbers listed in Labarrière et al. (2020)), it could be possible to investigate the cross-subject reliability of ML algorithms. It opened the opportunity to investigate whether training machine learning algorithms on non-amputated individuals could be used to detect physical activity in ILLAs; similar research has been conducted in the past with stroke patients with mixed results (O'Brien et al., 2017; Vageskar, 2017). If successful, this would have significant clinical applications: it would signify that the training data for a machine learning algorithm would not need to be collected from vulnerable population, which can be dangerous, time consuming and expensive (Yin, Yang, and Pan, 2008).

4.3.3 The Sensor Configuration

To re-iterate the scope of the review, wearable sensors were the only considered method of sensory deployment, as on-object and video monitoring methods of HAR can ethically only be evaluated in laboratory conditions, but as just established in Section 4.3.2, laboratories were inaccessible due to covid-19 restrictions. Section 4.2.2 listed a wide variety of wearable sensors that have been used for application of HAR, including the various ways they can be embedded in 'smart' technology such that they do not have to be a standalone module. To determine the most appropriate sensory mechanism, it was decided to use the most common and reliable method described in the literature, which fortunately had previously been researched by Demrozi et al. (2020) and is summarized in Figs. 4.16-4.17. Accelerometers are overwhelmingly the most popular sensor choice in HAR, and show consistently high performance in terms of accuracy across almost 150 independent studies. Recall that the list of outcomes that were desirable for clinical activity monitoring were identified as follows: stepping activity on slopes, stairs, camber (of a road), uneven terrain, and indoor/outdoor stepping, as well as prosthetic wear, cadence variations, walking speed variations and going cycling. In essence, the outcomes are centralized on walking activities, with the exception of cycling and prosthetic wear. This chapter has demonstrated that walking activities (and cycling) are all capturable with an accelerometer, meanwhile prosthetic wear can be inferred from a lack of acceleration over an extended period of time. This showed that an accelerometer, even standalone, was a suitable device for recognizing the identified outcomes.

In Sections 4.2.2.1.2 & 4.2.2.1.3: it was argued that physiological and environmental sensors were rarely used in activity monitoring due to their fundamental flaws, and the findings of Demrozi et al. (2020) show that they are typically relegated to niche uses in HAR. Though EMG and heart-rate sensors are listed as the most accurate sensors, the number of studies that have used them are astronomically smaller than accelerometers, thus their validity as sensing devices have yet to be proven. It is evident from Vaizman, Ellis, and Lanckriet

(2017)'s study that environmental sensors do not necessarily improve activity recognition rates, but rather can assist with the annotation process. In their study, they used audio and GPS signals to help decipher the contextual environment where a physical activity was being performed, which in turn made the annotation process less prone to errors. Further evidence is shown in Md et al. (2016), which used heart rate and barometric sensors to "context filter" activities primarily measured through accelerometer and gyroscope sensors. However, even if these sensors were included for context filtering, one must consider that in a clinical environment, requiring a client to wear an array of sensors would probably be physically uncomfortable for the client. Additionally, requiring more sensors would significantly increase the level of compliance on the user-end (meaning more demands would be made of the participant to ensure they were using all devices correctly at all time), which would not translate well to clinical environments. It became the aim of the activity monitoring study that a minimal amount of sensors and wiring should be needed to perform HAR. Therefore, physiological and environmental sensors, as well as gyroscopes and magnetometers, could optionally be included so long as they were incorporated with the accelerometer in a singular embedded device.

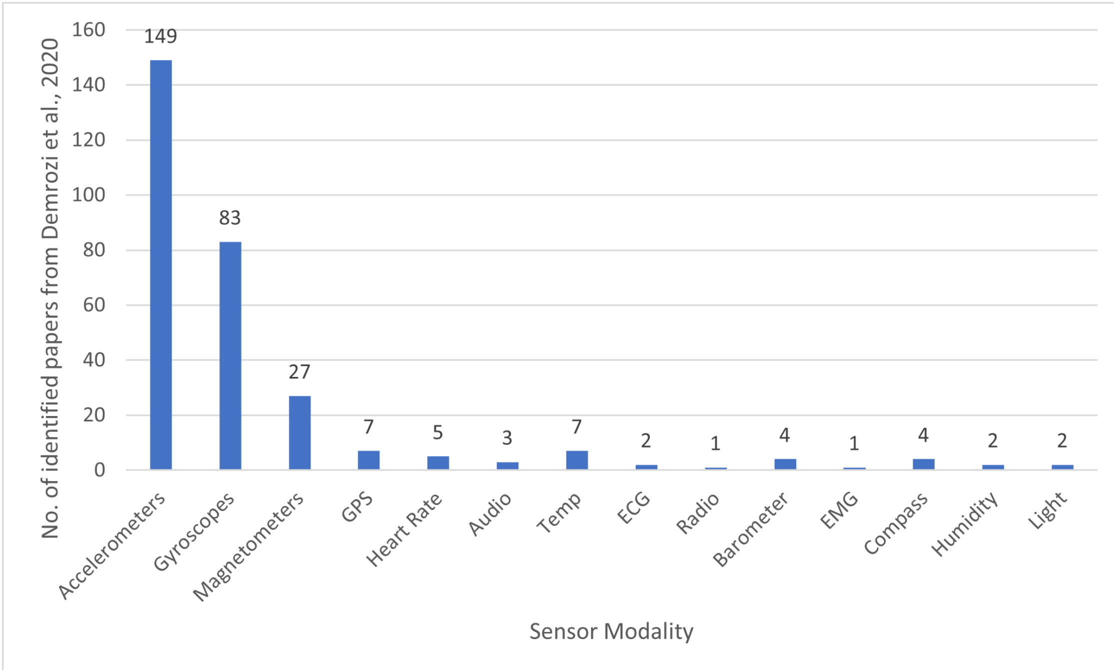


Figure 4.16 Barchart of most popular Sensors in HAR literature, as identified by Demrozi et al. (2020). Recreated from Demrozi et al. (2020)

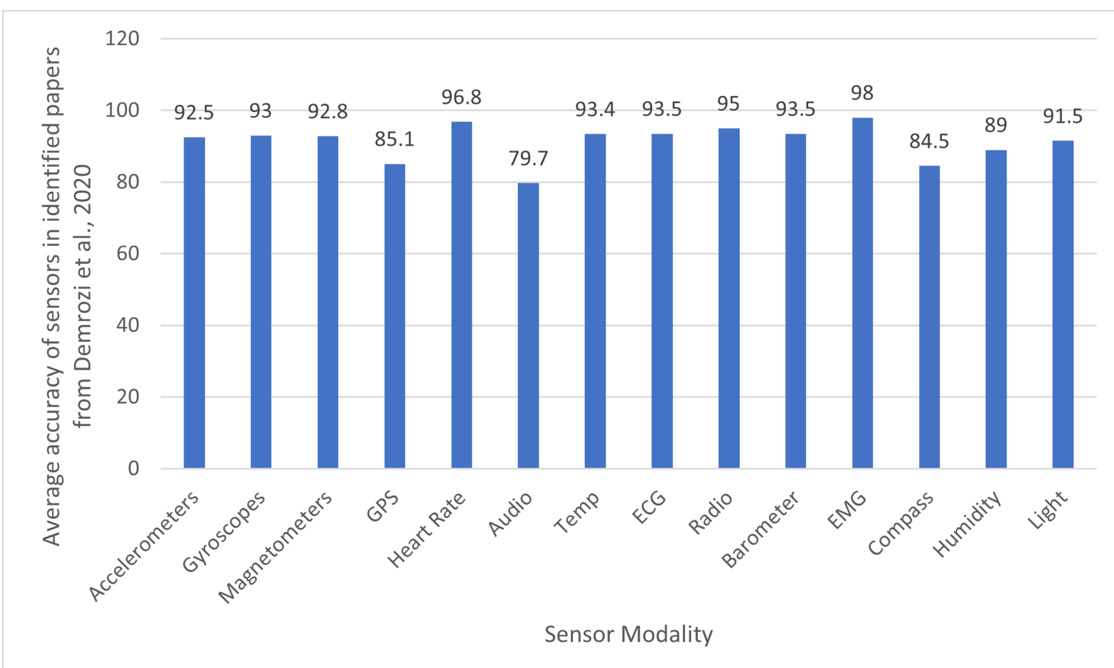


Figure 4.17 Barchart of the average accuracy of sensors in HAR literature, as identified by Demrozi et al. (2020). Recreated from Demrozi et al. (2020)

With primary method of sensing established (accelerometry), the next decision was to choose the most suitable accelerometer, or a multi-sensory device embedded with an accelerometer. Because it was impossible to directly oversee participants in the study, it was important that the wearable sensors used for the investigation had a long battery life such that they would endure operation for at least a week without needed maintenance from the user. When battery life of the wearable sensor is short, there will inevitably be less compliance from the participants, leading to the loss of large volumes of data (Sanders et al., 2018). From these considerations, the ActivPAL IMU was chosen as the sensory device. The ActivPAL has proven reliability in being able to distinguish between sitting, standing, and walking (Grant et al., 2006), and has demonstrated ability to measure step counts reliably in ILLAs (Buis et al., 2014). The ActivPAL's primary sensory instrument is a 12-bit resolution accelerometer, which operates at 20Hz sampling rate and is considered a good sampling frequency for HAR (Khan et al., 2016). Through dynamic data compression in the absence of movement, the ActivPAL is capable of comfortably operating over the course of a week without requiring charging; the long battery life of the ActivPAL also makes the device suitable for clinical monitoring, as short battery life in sensory instrumentation was one of the barriers to prosthetic monitoring identified by Chadwell et al. (2020). The device is also tamper-resistant (a key quality needed in a monitoring device as indicated by HCPs in Chapter 3.2.3), as recording can only be interrupted by connecting to specialised software. Another important factor to disclose was that the PhD was carried out with sponsorship from PAL Technologies, which meant their team members could provide the latest up-to-date hardware revisions and software and assist with giving technical advice relating to the ActivPAL. The ActivPAL unit additionally contains a magnetometer operating at 20Hz, and a thermometer and barometer which operated at roughly 1Hz each. These modalities are discussed further in Chapter 5.6.1. Another interesting detail about the ActivPAL is that it is paired with proprietary software (PAL Analysis, PAL Technologies, Glasgow UK) that uses readings from the ActivPAL to automatically process periods of sitting, standing and walking bouts and can also record step count. The application of this software to validate step count and bouts of movement or standing is explored in Chapter 8.

4.3.4 Choosing number and placement of Sensors

The key findings of the literature review of number and placement sensors in Section 4.2.3 was that there generally is not an optimal location for placement of a wearable sensor, however all wearable sensors should account for displacement of the sensor during movement as a potential confounding factor on sensor readings. When placed on the anterior thigh, a single ActivPAL can provide ample coverage of physical activity movement in the lower body. By including more ActivPALs, there is a risk of reducing user compliance. Additionally, it was desirable to try and keep the costs of the study low; by including more ActivPALs, this would increase the price of replicating the investigation. The ActivPAL by design has an optimised step-detection algorithm when the sensor is placed on the anterior thigh (Edwardson et al., 2017). This was also why the ActivPAL was favoured over a smartphone: by ensuring the location of the sensor's placement was consistent, there would not be a confounding disorientation factor in the ActivPAL sensor readings and would eliminate any need to create an algorithm to correct for sensor displacements. The ActivPAL is a small 4x2cm device can also be worn under clothing, which assists with compliance as participants are generally more likely to wear a sensor that is not noticeable (Bergmann and McGregor, 2011).

4.3.5 Feature Extraction

Section 4.2.4 detailed the various types of features that can be collected for wearable sensor data. While the features that can be extracted are mechanically simple (for example, the mean of a sample of data), it is generally impossible to determine beforehand whether those features will benefit the classification performance or not. Features that may work well in one HAR study may be completely irrelevant in another HAR study. The research by

Labarrière et al. (2020) contained an exhaustive list of features used in ILLA HAR and they are primarily a mix of common statistical features, suggesting there are currently no known features that can be employed specifically relating to the detection of ILLA physical activity.

The feature list for the investigation was built through a combination of consideration for popular features for HAR that were outlined in this chapter, and an exploratory data analysis using the ActivPAL sensor data in timeseries plots. From this chapter, the relevant features that were identified through the literature and Labarrière et al. (2020) were statistical features including the mean, median, variance, root mean square, norm, mean absolute deviation and the differentials of these signals, which appeared with abundant frequency. As there have been no investigations to dispel the effectiveness of these features within a HAR context, there was no compelling reason to not include them. These features are all non-specific time-series based and can be applied to any sensor. The exploratory data analysis involved assessment of data collected after the HAR study was completed, and is thus elaborated in Chapter 5.7.

When using deep learning neural networks, the need for feature engineering can be bypassed entirely, as networks like the CNN and LSTM can automatically make inferences about important features when given supervised (annotated) training data that can consist of just the raw timeseries data from the sensor. However, the "relevant" features that are chosen in deep learning are often far more abstract and may not be interpretable. For the purposes of understanding the data and discussion of important features, especially in comparison of non-amputated and ILLA populations, it would have been unfavourable to disregard the feature engineering process together. The application of deep learning networks for this human activity recognition study are further discussed in Section 4.3.7.1.

4.3.6 Feature Selection

It is important to recognize when feature selection or feature transformation should be chosen as the dimensionality reduction method. In the literature, there appears to be no papers which have considered the best dimensionality reduction approach relevant to ILLA participants. For the general population, neither approach has been shown to be consistently the stronger method; in Rasekh, Chen, and Lu (2014), LDA had much better classification rate over SFS approaches, whereas in Dehzangi and Sahu (2018) and Wang, Cang, and Yu (2016) a wrapper and hybrid filter-wrapper approach respectively outperformed PCA and kPCA. Throughout this chapter, the literature suggests that feature selection should be chosen when the individual features are of importance (i.e are certain features calculated relevant or irrelevant), whereas feature transformation may be preferable if the only consideration is the performance of the classifier system, as the relevant features are harder to interpret. Finally, it should be noted that feature selection and feature transformation are not mutually exclusive processes; hybrid approaches, such as those shown in Abo El-Maaty and Wassal (2018) and Wei et al. (2018) demonstrate that it is possible to first perform feature selection to remove redundant features, and then apply feature transformation to further compress the relevant features into lower dimensions. Both trials shown promising results, obtaining 95% accuracy with reasonably sized numbers (10-20) of PCA components. For the subsequent investigation, it was decided to investigate all three approaches: to apply feature selection and dimensionality reduction separately, and to apply feature selection followed by dimensionality reduction. For feature selection methods, chi-square, Relief-F and minimum redundancy Maximum Relevancy are chosen due their pre-programmed availability in Matlab's classification package. Matlab was the chosen environment for data processing, which is elaborated in Chapter 5.3. At first, only PCA was considered as the dimensionality reduction process due to its high computational efficiency and ease of interpretability (regarding understanding which features explain the systematic variance) compared to non-linear dimensionality reduction processes (Karamizadeh et al., 2013). Non-linear dimensionality reduction approaches like t-SNE and UMAP were initially not considered for implementation due to their output inconsistency created by their non-convex cost functions, but were eventually included in Chapter 7, in this chapter an unsupervised learning approach is taken (see Section 4.3.7) and in that analysis it no longer became important to understand which features were chosen to best represent the data in low dimensions, allowing for the incorporation of the non-linear processes.

4.3.7 Classification Approach

In Labarrière et al. (2020)'s systematic review of human activity and intent recognition studies, all of the studies conducted used supervised models or neural networks; none of the identified studies used an unsupervised model approach. A literature search in PubMed, Scopus, and Web of Science databases was then conducted using key search words: *"unsupervised", "amputation", "physical activity"* (and numerous synonyms for each key term), yet no studies were identified. Ariza Colpas et al. (2020) conducted a systematic review on unsupervised learning studies for HAR, and none of the included studies used ILLA participants. These findings all strongly indicate that an unsupervised learning study had never been conducted with an ILLA population. While unsupervised models typically tend to underperform in HAR contexts compared to supervised models (Ermes et al., 2008), the unstructured approach to unsupervised learning can potentially lead to the identification of new activities not defined by the researchers. In regard to the activity monitoring system discussed in Chapter 3, the unsupervised approach could allow the system to evolve and capture new types of activities. Recall that the number of ILLA participants recruited for interview sessions were low, and so it is likely these interviews did not encapsulate the full portfolio of activities for an ILLA. All unsupervised classification algorithms discussed in this review were considered for analysis, as all had shown great capability of clustering activities under extrinsic evaluation techniques. As it was felt it was necessary to have ground truth annotation of the dataset in order to give some concrete evidence of reliability of the algorithms, extrinsic evaluation of unsupervised clustering algorithms would be possible within the confines of this HAR study.

Supervised analysis of the dataset was also investigated; as previously mentioned, the number of HAR investigations conducted on an ILLA population in free-living conditions is very low, and none of the identified studies were conducted using only inertial wearable sensors, so there was still novelty in applying a more traditional supervised approach. The supervised classifiers chosen are those described in Section 4.2.6 - namely the SVM, kNN, RF, AB, BN and the LDA approach, the latter of which was discussed as a possible feature selection and classification approach. This is due to their preconfigured integration in Matlab's classifier app, where hyperparameter tuning can be carried out with minimal need for coding. Each of these classifiers were demonstrated in this review to outperform or underperform each other in a multitude of HAR studies, therefore it seemed probable that any one of these classifiers could perform the best in the constructed study. While there was opportunity to add complexity to the classifier design, for example by combining the classifiers into an ensemble, this would generate considerable increase in the research scope. For example, if creating an ensemble, this creates multiple experimental scenarios that would have to be addressed: what classifiers should be chosen? How would they integrate with each other? Would multiple combinations be tested? Would classification be varied with different structures of activity labels? These scenarios could not be completed to a satisfactory degree within the timeframe of the PhD, and as such only looks at preconfigured classifier models.

4.3.7.1 Considerations for Neural Networks

In addition to the supervised classifiers, neural networks were further considered as an alternative supervised classification method. As previously stated, they have been used before to classify walking activities carried out by ILLAs (Labarrière et al., 2020). By incorporating neural networks, this could provide an interesting discussion of feature engineering versus automated feature learning in regards to recognizing walking activities from wearable sensors. Neural networks would be implemented only on the condition that the collected dataset was of an appropriate size. It is a misconception that neural networks and deep learning require large dataset sizes (> 10000 samples for example) to adequately construct a classifier. If the number of samples collected in the investigation was comparable to numbers collected by other studies that have used deep learning for HAR, then a neural network approach would be deemed viable. Of the neural networks discussed in this review, only the LSTM classifier was considered for implementation, the reason why only LSTM was chosen was due to the general design complexity of neural network processes; it would consume a considerable amount of time to design multiple Neural Networks while also finding suitable hyperparameters. In Matlab deep network design interfaces, CNNs require input signals to be converted to an image format before being processed, and by considering the potential size of the dataset, this would require an exponential increase in the required data memory. The results of this are discussed further in Chapter 5.9.3.

4.4 Chapter Conclusion

This chapter has been a concise review of the concept of machine learning and human activity recognition. Through analysis of the literature, it was identified that in regard to the ILLA population, very few HAR studies are conducted in free-living conditions, and no research appears to have attempted to formulate a HAR system using an unsupervised model. These research gaps led to the formulation of a HAR study on ILLAs which will be described in detail in the next chapter. In regard to the activity monitoring system, a robust HAR classifier built with a single low-cost wearable sensor on the data of ILLAs could be utilised by healthcare professionals to monitor their activity.

Chapter Five

Collection of Detailed Physical Activity Data from Non-Amputated Individuals and Individuals with Lower Limb Amputation in Free Living Settings

5.1 Introduction

In Chapter 4, it was established that there are gaps in the research field relating to the HAR of ILLAs in free-living conditions. By conducting an investigation and acquiring a robust HAR system in these conditions, this research could lay the groundwork for a low-cost activity monitoring system that can capture relevant physical activities for healthcare professionals to analyse. In this chapter, the challenges of capturing the data in an environment where there were strict limitations due to the coronavirus epidemic are discussed. The data captured by the wearable sensor, the necessary pre-processing procedures and the quality of data gained from the investigation will also be analysed in detail.

5.2 Formation of the Data Collection Study

In order to conduct the investigation, ethical approval had to be sought from Strathclyde University. Ethical approval was sought separately for the two main populations of the investigation: ILLAs and non-amputated individuals with no known gait impairment. Ethical approval was granted for the non-amputated population (reference BioMed/2020/293) and for the ILLA population (reference UEC20/55).

University ethics dictated that under coronavirus restrictions, physical contact between the researchers and the participants would be minimized. This meant that face-to-face interactions were limited to delivering and returning equipment for the study, thus in order to acquire Ground Truth Annotation (GTA) for our data, the participant would need to be responsible for recording themselves.

5.2.1 Acquiring the Ground Truth

The GTA process is when raw data points are given labels (in this context, a physical activity) which indicates the true identity (the "ground truth") of the data point. As the ground truth must be determined through human inspection or analytical processes, the ground truth is not always the "absolute truth", meaning the activity that has actually been carried out. In supervised machine learning scenarios, labelled data is a necessity as it trains a classifier to associate the values of features with classes, in this context activities. GTA also provides a method of evaluating the classifier's performance. There are various means through which a ground truth can be obtained, and each method has their strengths and weakness as discussed in the "Data Collection" section of Chapter 4.2.1.2. The final method chosen for GTA was to utilise a body-worn camera video recording set-up. Other methods simply could not provide accurate timestamps to detect activities. While there is risk of camera obfuscation, by instructing participants to wear a hands-free body camera worn over clothes, the risk of loss of data would be minimized. Unfortunately, using a camera came with some strict limitations from the ethics committees; it was not possible to allow recordings to be carried out in the participant's homes, and only recordings of public areas were allowed. To adapt to this limitation, the study was structured such that data collection would only take place when the participant went for a walk in their local vicinity. To protect visual identities of the public, the camera also had to be angled approximately 45 degrees downwards from the horizontal. This would have some impact on the difficulty of annotation, as will be discussed in Section 5.5.

5.2.2 Acquiring GPS data

From Chapter 3.2, one of the key activities to be established by an activity monitoring system is the identification of uphill and downhill movements. In a laboratory environment, sloping movement can be easily distinguished through the use of ramps. In a free-living environment however, slopes are much more difficult to define and measure due to the sporadic nature of terrain in external environments. In order to keep costs low, slope information was tracked using elevation data from the fitness-tracking Strava™ application (Strava Inc., CA USA) installed on a second-hand iPhone® 6 (Apple Inc., CA, USA) device. If the operating smartphone contains a barometer or altimeter, then Strava's elevation readings are directly extracted from the smartphone (Meg, 2021), hence the accuracy and precision of the Elevation readings were determined by the smartphone's sensors. The iPhone 6 contains a BMP280 barometer (Bosch, Gerlingen, Germany) which can measure changes in altitude with a precision of $\pm 0.12hPa$ (or $\pm 1m$ equivalent), operating at a sample frequency of approximately 1Hz (Bosch, n.d.; Hintz, Vedel, and Kaas, 2019). While the elevation precision of the sensor was poor, this was not a large concern as the exact elevation reading was not necessary to determine the presence of a slope, as will be elaborated in Section 5.5.1. Moreover, the altimeter included in the iPhone 6 has been found to have the highest elevation accuracies among consumer grade GPS devices, especially when using Strava (Barberi, 2017), so this appeared to be the most cost-effective method of obtaining accurate GPS readings.

5.2.3 Study Requirements

Of notable inclusion criteria, participants from both demographics were required to be able to carry out physical activity for a sustained period of time. In order to try and accommodate for older participants, it was determined that over the course of a week, (the approximate maximum operating period of the ActivPAL) participants should be able to collect 20 minutes of walking data per day, totalling 140 minutes of data per participant. The low threshold of 20 minutes was chosen to ensure that all ILLAs, whom statistically have less fitness and endurance than the general population (Chin et al., 2002), would avoid fatigue. All participants were given flexibility over the recording period; for instance, they could complete recording in 2x70 minute sessions provided they were comfortable with doing so. To adhere to the free-living component of the study, participants were not instructed to take a particular route prescribed by the researcher; each participant was given the opportunity for the researcher to recommend a particular route devised from Google Maps™ (Google, CA, USA), however all participants declined the offer. As terrain variation was one of the identified ideal outcome measures of an activity monitoring system (from Chapter 3.2), participants were encouraged to walk on as many a variety of terrains as they were comfortable with, including sloping variations of terrains and up and down stairs. Given that the target population was ILLAs and recording was to be held in exterior environments during the Autumnal and Winter seasons, participants also had to be instructed to not perform vigorous activities (for example jogging) for risk of slipping and falling during wet or icy weather. As technology was an important part of complying with the study, all participants who consented to participate were given a brief survey through e-mail communication to ensure they would be able to operate all technological components without issue.

In summation, the study would be conducted with participants recording themselves walking only in external environments for approximately 140 minutes over the course of a week. There would be 3 recording devices used: the ActivPAL, which would capture IMU data, a body-worn camera, to provide GTA relating to activities and terrains, and a Strava recording running on an iPhone 6 device to capture elevation data. Participants were instructed to place the iPhone in a pocket or jacket away from the ActivPAL worn on the thigh, such that it would not interfere with the ActivPAL's magnetometer sensor. An example of the recording set-up is shown in Fig. 5.1. As the ActivPAL and the smartphone would be worn under clothing, their approximate locations are illustrated. The smartphone's placement was arbitrary, so long as it was not placed in a pocket next to the ActivPAL device, which could potentially create noisy accelerations if the two devices bumped into each other.

Figure 5.1 Photograph of recording setup for the data collection study (AP = ActivPAL).

5.2.4 Recruitment of Participants

As university campuses were closed throughout 2020 due to Covid restrictions and there was no foreseeable reopening in 2021, it was not feasible to recruit on campus. As a result, most participants were recruited through convenience sampling; non-amputated participants were recruited by advertising on the author's social media. ILLA participants were recruited through a combination of contacting participants from the previous interviews of Chapter 3 (who had given consent to be contacted again for future studies) and through a virtual poster displayed on a local amputee charity's social media page. This led to some volunteering and location bias in the participant demographics. All participants had to be at least 18 years of age, be comfortable performing moderate vigorous activity and not be at risk of life-threatening conditions if infected with the coronavirus. The ILLA volunteers were further required to be able to ambulate with a prosthesis without the use of walking aids and not have any comorbidities or musculoskeletal conditions which could potentially impact on their ability to ambulate for sustained periods of time.

5.2.4.1 Participant Characteristics

A total of 8 non-amputated participants and 4 participants with lower limb amputation were recruited for the study. Their primary characteristics are listed in tables 5.1-5.2. From this point onwards, specific subjects are referred to using a letter-number rule. For example, non-amputated subject #8 is referred to simply as "H8", and amputee subject #2 is referred to as "A2". All participants who consented to participate in the study were able to carry out their recording successfully, with zero dropouts. However, it should be noted that two of the recordings from subject H1, and one recording each from subjects H4 and H5 had to be discarded from analysis due to improper camera positioning or recording at nighttime, making video data poor and analysis of those recordings subsequently infeasible. The demographics match in terms of height and weight, however there were large gender biases in both demographics with the vast majority of participants being male. Additionally, there was a noticeable age bias between non-amputee and amputee demographics; most of the non-amputated volunteers were in their mid-20s, while the mean age of the ILLAs is approximately 50. While this is undoubtedly could have had adverse impact on machine learning testing accuracies, for research purposes it would be useful to determine if training data on a younger non-amputated population could still result in the detection of activity for an older population with gait impairments. Despite the relatively small sample size of amputee subjects, there was some interesting variation in the amputation time with 2 long-term experienced amputees and 2 comparatively inexperienced amputees, and there was a bilateral amputee to provide a comparison point to unilateral amputees.

Table 5.1 Characteristics of the non-amputated participants

Subject	Height (m)	Weight (kg)	Age (years)	Gender
Non-amputated Subject #1	1.80	84	24	Male
Non-amputated Subject #2	1.65	63	51	Female
Non-amputated Subject #3	1.62	65	18	Female
Non-amputated Subject #4	1.97	99	25	Male
Non-amputated Subject #5	1.92	102	25	Male
Non-amputated Subject #6	1.83	89	24	Male
Non-amputated Subject #7	1.84	88	25	Male
Non-amputated Subject #8	1.78	98	25	Male

Table 5.2 Characteristics of the ILLA participants

Subject	Height (m)	Weight (kg)	Age (years)	Gender	Type of Amputation	Amputation Time (years)
Amputee Subject #1	1.79	95	55	Male	Unilateral transtibial	3
Amputee Subject #2	1.70	86	57	Male	Unilateral transtibial	32
Amputee Subject #3	1.72	110	40	Male	Unilateral transtibial	33
Amputee Subject #4	1.52	61	48	Female	Bilateral transtibial	4

5.3 Computing Environment

All data processing in this chapter and all chapters following are performed in the Matlab environment (Matlab r2021a, Mathworks, MA USA). Matlab was chosen due its ease of use - all function packages can be downloaded from a centralized secure server - and was a useful tool for debugging via the variable explorer and informative documentation for each of its functions. Matlab also allows for integration of community-built packages, which can be reviewed by other users to check whether the implementation of the package works as intended. All calculations were locally performed on the author's personal computer; the processor used for calculations was an i5-8600 Intel® CPU with 3.10GHz clock speed and had 16GB of RAM available.

5.4 Data Synchronization

Synchronization of the ActivPAL, camera and Strava timestamps were necessary in order to ensure that the annotated labels would be associated to the correct timestamps in the ActivPAL sensory data, and to ensure that the uphill and downhill annotation data would be accurate. Each source obtains a "time" using a different process: the camera time is manually set using in-built software and uses an internal clock to track the time. The Strava GPS timestamps are acquired directly from the operating smartphone, and the ActivPAL's internal clock is synchronized to the computer which initiates the device's recording process. Through some experimentation it was discovered that there was significant drifting in both the camera and the ActivPAL's internal clocks. By using points in video recordings where the participant was standing still and comparing this to acceleration readings in the raw ActivPAL data, it was determined that the ActivPAL's internal clock drifted at a constant rate of 16 seconds per day relative to the time of the camera's clock. This information was used to adjust the correct time of the ActivPAL's internal clock. Further, as the drift factor was approximately 0.5 seconds per hour (relative to the camera time), longer recording sessions adjusted the time of the ActivPAL an additional 0.5 seconds for each additional hour of recording. In order to correct for the camera's drifting factor relative to the GPS timestamps, the participants were instructed to use the camera to record the clock on the iPhone's display screen at the start of each recording, and this determined that the camera drift factor relative to the GPS time was approximately 3 seconds per day.

5.5 Data Annotation

After data had been collected from all participants and synchronized, the first stage of data analysis was to perform data annotation. Annotation was performed with Visual object Tagging Tool (VoTT) software (Microsoft, WA, USA). While the software's primary function is for image recognition training, VoTT contains the ability to playback any video in real time and quickly annotate scenes. It also contains a useful function which allows the user to cycle through each annotated scene, such that it allows easy review of the annotated tags for each video. The program can export the timestamps of each annotation in a ".csv" format, which can then be used to associate sensory data with the labels in other software. Fig. 5.2 shows an example of the VoTT annotation process for one user.

Approximately 1,680 minutes of video data was collected (12 participants * 140 minutes of data). The time required to complete annotation for a single video was roughly equal to the length of the video itself; when there is a change in activity or terrain - for example going from a grassy uphill terrain to a concrete uphill terrain - the video had to be paused to assign the label. However, the video playback speed could also be fast-forwarded up to five times, which was advantageous when it was clear there was no change in terrain, which roughly balanced the time of annotation to be equal to that of the recorded time. To mitigate fatigue and error in the labelling process, the author would take a 10-minute break after completing an hour of annotation. While multiple rounds of annotation process were considered (meaning

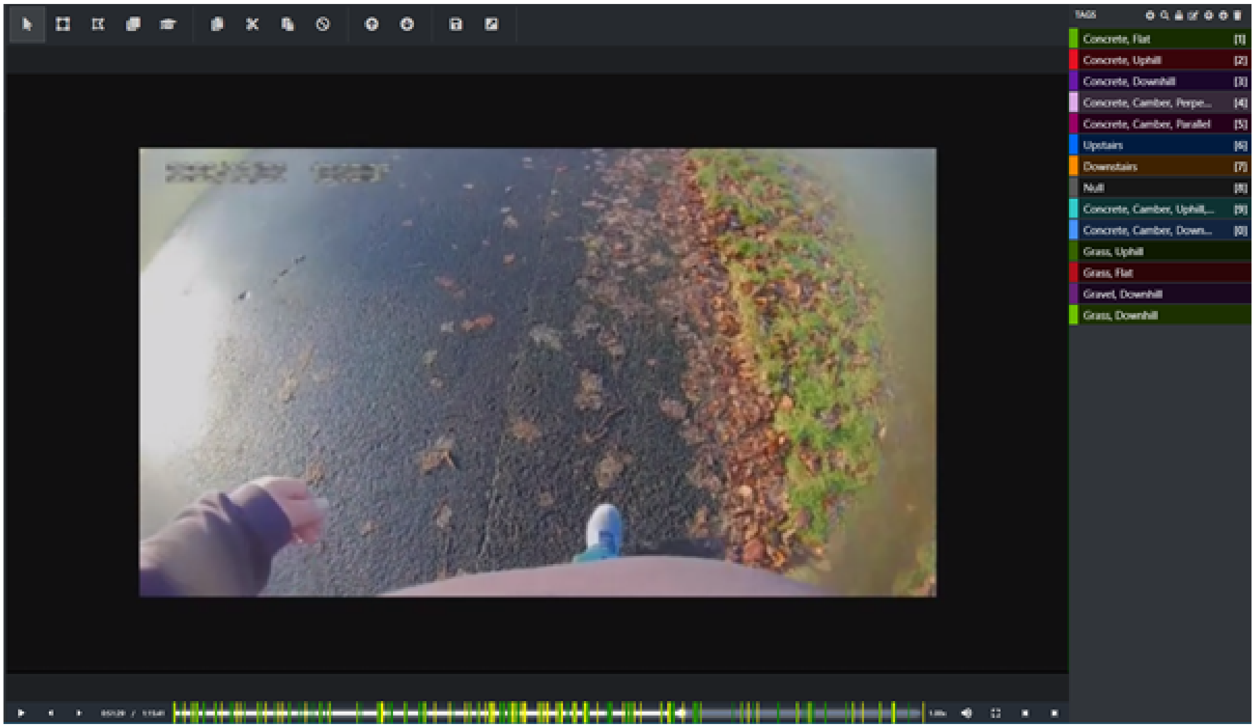


Figure 5.2 Screenshot of VoTT user interface

perform annotation process multiple times then obtain an 'average' annotation label), the full round of annotation required 4 full-time working days to complete, making it impractical and fatiguing to repeat the process again. The impact of only performing a single round of annotation is explored in Chapter 6.4.4, where it was speculated there were potentially numerous instances where the ground truth annotation acquired in this process did not necessarily match the absolute truth of the activity label.

Activities were labelled first by the type of terrain that was being traversed (e.g concrete, grass, sand etc.) and then by the condition of the terrain (flat, uphill, or downhill). On certain roads that physically resemble the example in Fig. 5.3, additional information was added whether the subject was crossing the road ("Concrete, Camber, Perpendicular") or walking in parallel to the direction of the road ("Concrete, Camber, Parallel") as both activities can make slight alterations of the gait: in perpendicular camber movement, there is a slight rise followed by a slight fall in elevation as they cross the "crown" of the road, while in parallel camber movement one leg is positioned slightly higher than the other.

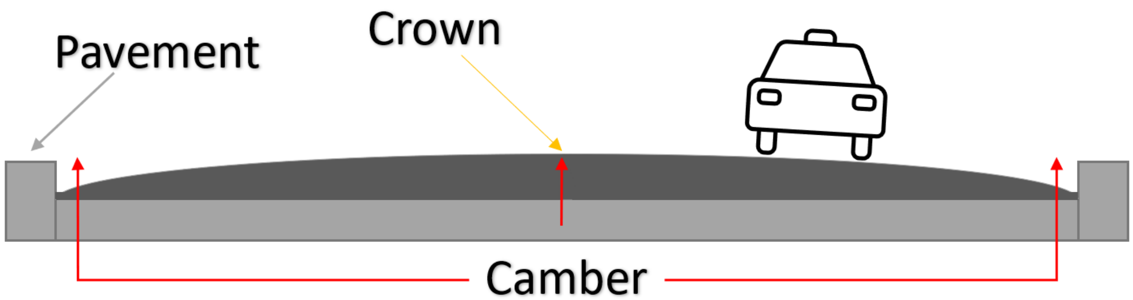


Figure 5.3 Diagram of the key components of a camber road, recreated from SketchUP3DConstruction (n.d.).

In addition to labelling the timeseries data with walking activities, the "Null" tag was used in a variety of situations whenever there was an event that could lead to misclassification in the HAR algorithms. There were a handful of conditions in which scenes would be given a null tag:

1. During any stationary activity. Given the ActivPAL's reliability in being able to distinguish between sitting, standing, and moving through its proprietary VANE algorithm (Grant et al., 2006), it was decided to exclude all stationary movement from HAR recognition. In Chapter 8.3, there is some discussion on how the ActivPAL's VANE algorithm could be combined with machine learning to automatically discard stationary activities from machine learning analysis.

2. Any recorded video that was taken in or near the participant's homes.

3. Anytime the camera footage was obscured for a sustained period of time.

4. Any niche instances of movement (e.g walking backwards, shuffle of feet) as there would be no feasible way of training and recognizing these activities given the small sample size.

5. Whenever the terrain could not be confidently identified by the researcher

6. Slope angles within an ambiguous threshold range (see Section 5.5.1).

5.5.1 Annotation of Slopes

Slopes were notably difficult to annotate, in part this was due to the required downward angle of the camera making uphill and flat sections indistinguishable in certain recordings. Thus, in order to identify uphill and downhill segments of a walk, the elevation data from Strava was utilised. The angle of slope being traversed as a function of time, $\Theta(t)$, was calculated by approximating the slope as a straight line (see Fig. 5.4), and then calculating the inverse tangent of the elevation change, $\Delta Y(t)$, divided by the distance change, $\Delta X(t)$:

$$\Theta(t) = arctan(\frac{\Delta Y(t)}{\Delta X(t)}) \tag{5.1}$$

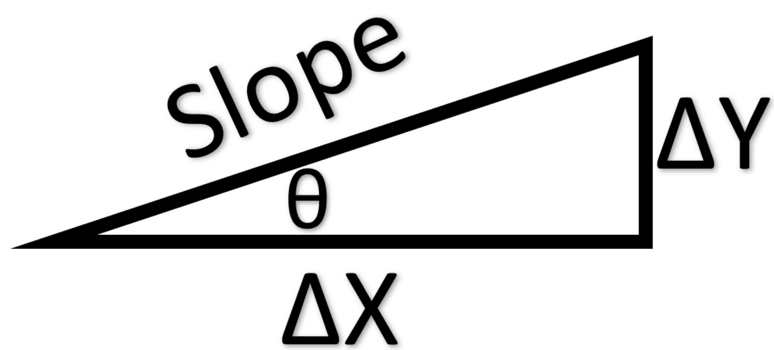


Figure 5.4 Diagram of slope angle approximation

All angle calculations were performed using a MATLAB script and Strava GPS data. ΔY was obtained from elevation readings, while ΔX was calculated using WGS84 ellipsoid distance of latitude and longitude co-ordinates. Since elevation readings were dependent on the accuracy of GPS readings, whose accuracy can fluctuate depending on the strength of the GPS signal, the angular values were given a smoothing average over a heuristically determined 10 samples, which gave a reasonable trade-off between "stability" of the angular values plotted over time and the peak magnitude of the angle readings. A failsafe counter-measure was put in place where if there were more than 15 seconds between two successive Strava readings, the angle would not be calculated and all unclear movement within that timeframe would be given a null label. An exemplary angle plot is demonstrated in Fig. 5.5.

As annotation of the video progressed, the annotator would periodically check for whenever the thresholds in the angular data were crossed, using the intersecting "*" points for precise guidance on when the threshold cross occurred. There were four key angular thresholds in the annotation process: a slope would be determined as uphill when the angle of the slope exceeded +2.9°, or -2.9° for downhill. This threshold was referenced from the American Disability Association's minimum slope angle requirement for construction of a sloped handrail in a public access building (DoJ, 2010) [1]. Due to the imprecision of elevation and ellipsoid distance readings, an "inconfident" threshold was also constructed between +1.45 to +2.9° and -1.45 to -2.9°. Any data that fell within this threshold would be given a null label, the exception being when a participant was traversing stairs. This led to unavoidable loss of data, but as it could not be confidently ascertained whether readings within this threshold were uphill, downhill, or flat, it would have led to numerous mislabelling instances.

5.5.2 Activities Collected

A total of 24 unique activities were obtained from data analysis. Table 5.3 compiles all activities collected from the study from a 40 sample or 2 second segmentation window, the choice of sampling length is explained in Section 5.6.3. Null labels were excluded from analysis. These activities use "full terrain resolution", meaning that as well as including the slope of the activity (flat, uphill, or downhill), the type of terrain that was crossed was also included. In later chapters, these labels undergo a "label consolidation" process, where similar activities are grouped together as a single, consolidated label. An example of the consolidation process for flat terrain activities is presented in Fig. 5.6. This hierarchy also applied to uphill and downhill activities. Stair data was considered to be "terrain-independent", as all steps traversed in outdoor conditions can be said to be formed from concrete or stone, and so all stair activities did not specify terrain.

[1] Ramp standards differ between countries, but given physical disability is a universal condition, the choice of standards is arbitrary

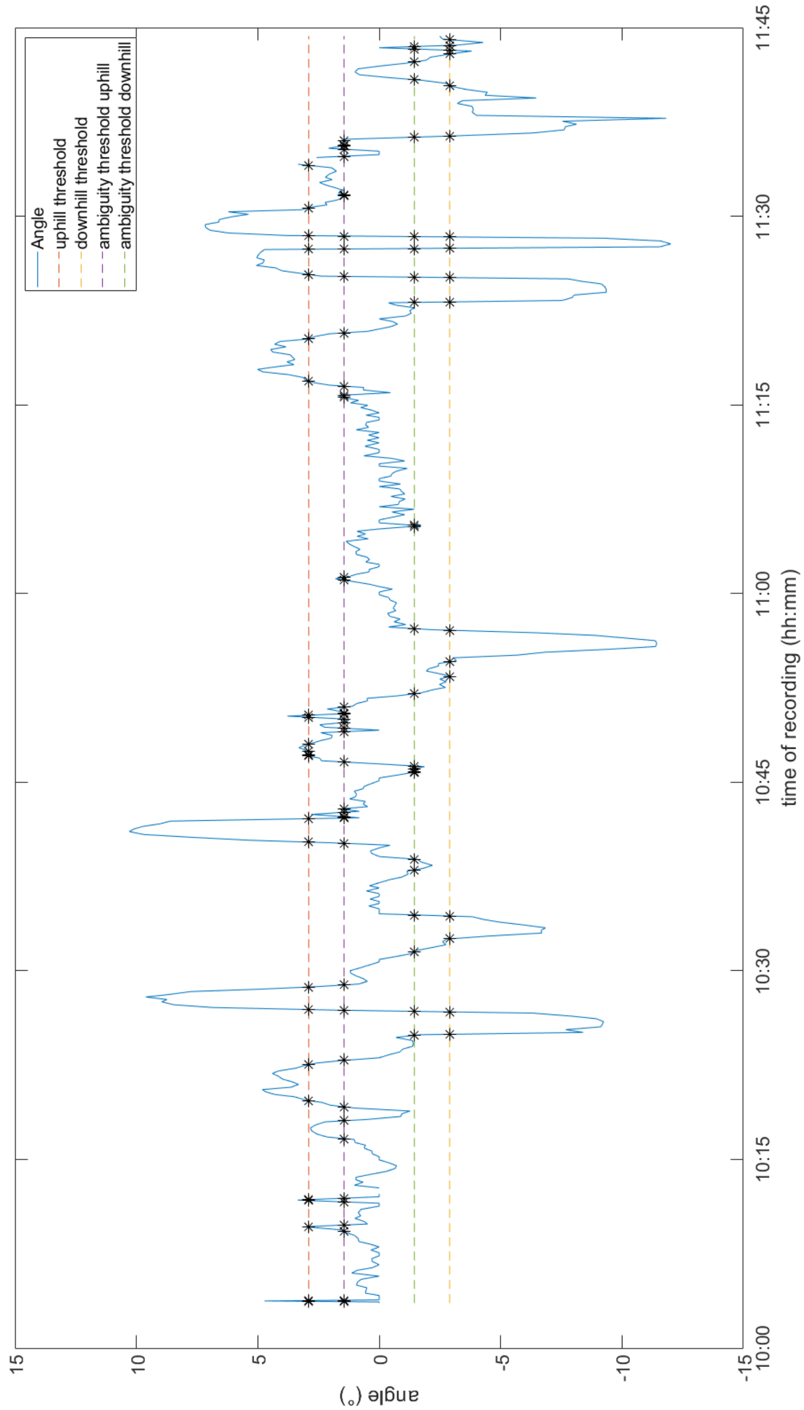


Figure 5.5 Plot of slope angle against time for one recording of a participant.

Table 5.3 Total sample counts of each activity with 40 sample window length

Activity Label	Total Sample Count
Concrete, Camber, Downhill, Parallel	572
Concrete, Camber, Downhill, Perpendicular	93
Concrete, Camber, Parallel	3923
Concrete, Camber, Perpendicular	877
Concrete, Camber, Uphill, Parallel	540
Concrete, Camber, Uphill, Perpendicular	93
Concrete, Downhill	2549
Concrete, Flat	12200
Concrete, Uphill	2427
Downstairs	656
Grass, Downhill	151
Grass, Flat	1346
Grass, Uphill	78
Gravel, Downhill	46
Gravel, Flat	32
Gravel, Uphill	3
Red Ash, Flat	26
Sand, Downhill	12
Sand, Flat	232
Sand, Uphill	31
Stone, Downhill	80
Stone, Flat	496
Stone, Uphill	176

Upstairs	691

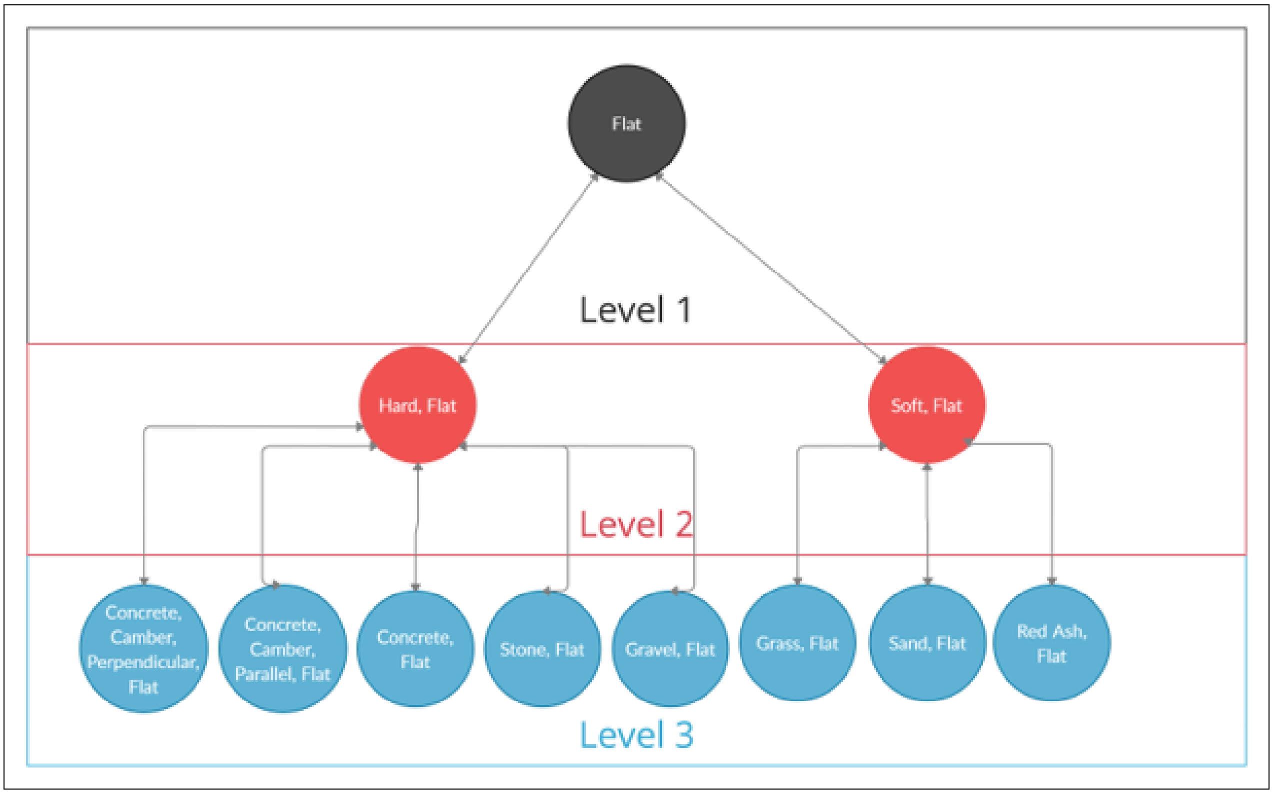


Figure 5.6 Label hierarchy for "flat" terrain activities

5.6 ActivPAL Sensor Usage and Pre-Processing

5.6.1 Sensor Usage

As previously discussed in Chapter 4.3.3, the only sensor used for HAR was a thigh worn ActivPAL, pictured in Fig. 5.7, which comes equipped with an accelerometer, magnetometer, barometer, and thermometer. The Accelerometer and barometer both operated at 20Hz, while the barometer and thermometer operated at approximately 1Hz. Due to their low sampling frequencies, information from the barometer and thermometer was discarded from further analysis. The accelerometer data from the ActivPAL was used for HAR. As previously established in Chapter 4.3.3 by Grant et al. (2006), the ActivPAL's accelerometer is considered a gold standard device for distinguishing sedentary and active behaviour.

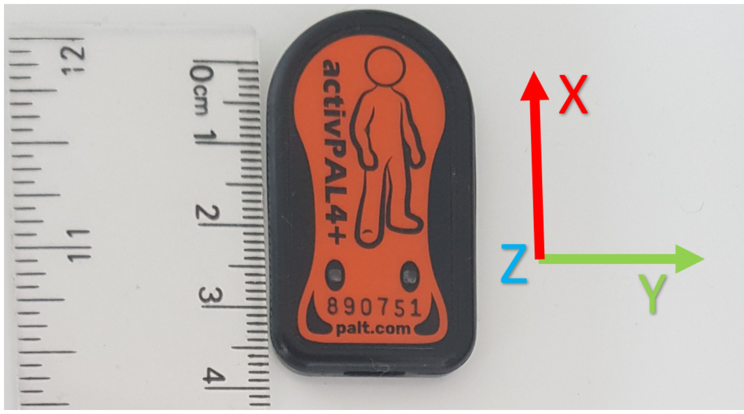


Figure 5.7 Photograph of the ActivPAL unit with ruler for scale. The adjacent axes represent the directions of accelerations when worn on the anterior thigh

The accelerometer readings were calibrated through PAL Technologies' proprietary "STAC" algorithm. Provided that there is a sufficient amount of data collected from the ActivPAL recording, the STAC algorithm will calibrate accelerometer readings with a precision of 95% or greater at ±0.02g, as Fig. 5.8 demonstrates. The left sphere represents sample readings taken by the accelerometer, with precision represented in a colour heatmap. High precision readings are indicated by green voxels, while lower precision readings are indicated by red voxels. On the right post-calibration sphere, all readings are in green indicating that all accelerometer readings have high precision. In a clinical environment, these devices would be calibrated and tested in-house thus not require this procedure.

The ActivPAL's magnetometer was at first intended to be used alongside the accelerometer. Offering an additional 3 DOFs and the ability to acquire North-East-Down (NED) orientation data via accelerometer and magnetometer sensor fusion could have provided useful features to distinguish physical activities (Silva, Paiva, and Carvalho, 2021). Unfortunately, the magnetometer data suffered from technical issues. During some recordings, the magnetometer readings underwent saturation; Fig. 5.9 demonstrates an example of one

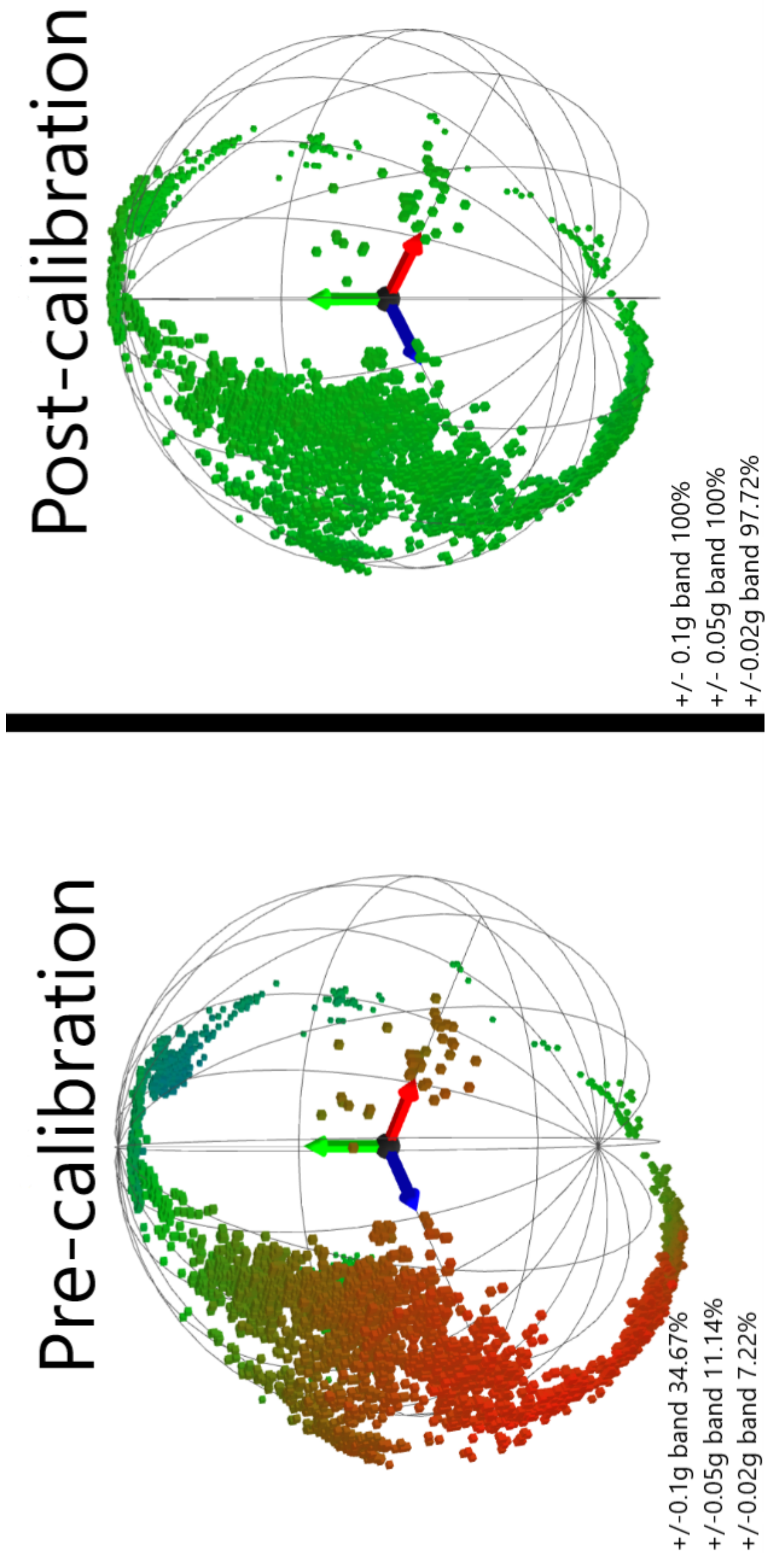


Figure 5.8 (Caption on next page)

Figure 5.8 ActivPAL reading precision pre-calibration and post-calibration, represented by colour-mapped voxel spheres. Red voxels in a particular direction indicate the accelerometer readings in that region are imprecise, while green voxels indicate higher precision bands of $\pm$ 0.1g or smaller

recording of a participant where the saturated readings were consistent throughout the entire duration of the recording, while Fig. 5.10 shows a different recording from the same participant where the magnetometer readings are not saturated. While it would be reasonable to discard such readings if they were only present for small periods at a time, discarding entire recording sessions was undesirable. When discussing this issue with the manufacturers of the ActivPAL, PAL Technologies, they warned that because the magnetometer implementation was still in a prototype phase, the magnetometer readings were not calibrated, and speculated that the saturated readings resulted from an incorrectly calibrated magnetometer, creating unstable gain in magnitude. At the time, they were working on creating an automated calibration algorithm, equivalent to the STAC algorithm for their accelerometer, but unfortunately this was not completed within the timeframe of the PhD. For these reasons, the magnetometer readings were ultimately discarded from the main analysis. Thus, only the accelerometer readings from the ActivPAL were utilised for HAR. However, the stable magnetometer readings were later used in an experiment detailed in Chapter 8.4.

5.6.2 Filtering and Denoising

When handling sensor data, a common technique is to apply a frequency-based filter to remove unwanted artifacts and preserve only the components of interest, in this context, this is the thigh movement of a human during walking. The configuration chosen for filtering the ActivPAL accelerometer was a band-pass filter with a lower cut-off frequency of 0.3Hz and an upper cut-off frequency of 4Hz. The lower cut-off was designed to remove the gravitational constant (Van Hees et al., 2013), but does not remove the effect of gravity altogether from the signal, particularly during high frequency movements (Chaurasia and Reddy, 2018). Consis-

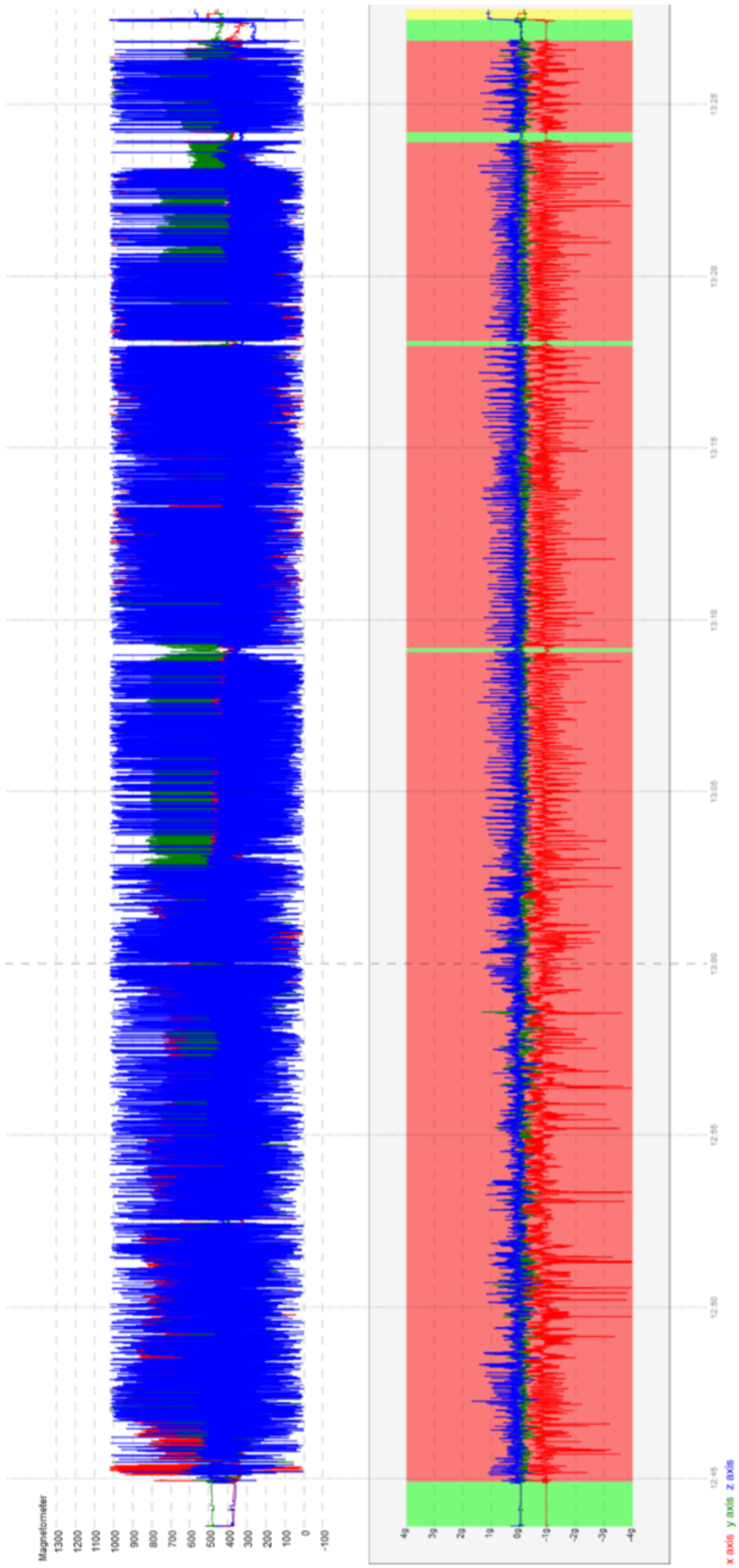


Figure 5.9 (Caption on next page)

Figure 5.9 Top: Unstable magnetometer readings of the ActivPAL recorded by participant A4. Bottom: Corresponding accelerometer data.

tent and complete removal of the effect of gravity would require installation of a gyroscope, whose readings are unaffected by gravitational forces (Nistler and Selekwa, 2011). However, this was not practical to do in a short time frame and could impact on the form factor and cost of the ActivPAL device. As all considered activities fall within the low frequency range (0.1-2Hz), the effect of gravity in the signal would have been overall negligible. The upper cut-off frequency meanwhile helps to remove high frequency noise generated through muscle movements in the thigh (Frølich and Dowding, 2018). As the ActivPAL is directly adhered to the skin, these forces would very much be present in the ActivPAL's readings. The chosen cut-off frequency provided a good balance between removal of unwanted noise and distortion of the original signal's values. Fig. 5.11 demonstrates a sample of the X-axis accleration (which is primarily the axis that is most influenced by gravity) from one non-amputated participant when no filtering, highpass filtering and bandpass filtering is applied. Noticeably, the highpass and no filter plots are almost identical save for the removal of the constant (making the mean acceleration approximately zero), while the bandpass filter removes some of the high frequency noise.

The bandpass signal was further analysed to determine whether or not additional denoising operations were required. De-noising conceptually has the same principles as filtering, as it is the process of removing unwanted (noisy) signals from the data. In HAR problems, de-noising is not always performed due to the risk of removing or distorting informative components of the signal. Erdaş et al. (2016) found no significant impact on HAR performance regardless of the type of filter applied. Some investigations have made a point to not carry out any de-noising and are still able to achieve respectable HAR performance (Ravi et al., 2005; Bayat, Pomplun, and Tran, 2014). To determine the viability of denoising techniques, random 10 second samples of accelerometer data in the X axis (approximately representing

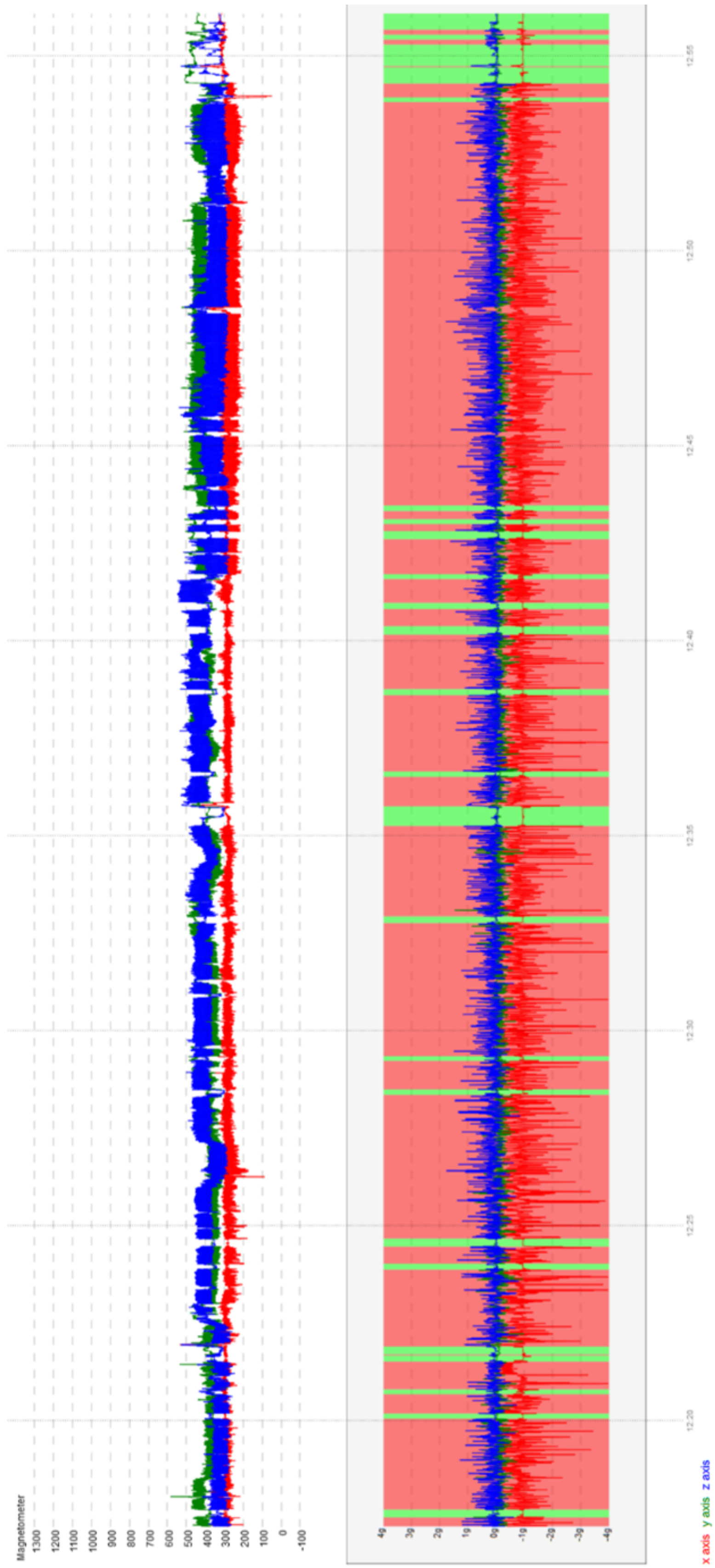


Figure 5.10 (Caption on next page)

Figure 5.10 Top: Stable magnetometer readings from participant A4, recorded on a separate day from Fig. 5.9. Bottom: Corresponding accelerometer data.

vertical accelerations when worn on the thigh) from one non-amputated participant with no amputation were extracted and analysed in Matlab's Signal Analyzer companion application. A moving average filter, Wavelet Denoiser and a Total Variation (TV) denoiser were all trialled after heuristically obtaining suitable parameters for each filtering approach, and these example plots can be seen in Fig. 5.12. From various samples that were explored, the wavelet denoised and smoothing average signals appeared to have minimal differences compared to standard application of a bandpass filter. TV Denoising distorted the signal by "squaring" the peaks, making accelerations appear unnatural. Thus, it was deemed unnecessary to apply additional denoising. To gain valuable qualitative assessment of the signal, a comparative paper with a similar methodological set-up would have been beneficial, but unfortunately to the author's knowledge no paper has used the ActivPAL's accelerometer data for human activity recognition. This is evidenced in Edwardson et al. (2017)'s review of ActivPAL studies, in which no included study explicitly uses the accelerometer data for analysis; researchers primarily utilize the ActivPAL's proprietary VANE algorithm to automatically process sitting, standing and walking behaviour. The closest comparitive study is the work by Huan et al. (2019), who utilised a smartphone accelerometer operating at 100Hz worn in a front trouser pocket, a location functionally identical to the ActivPAL's placement, to gather acceleration signals for gait analysis and extraction of gait features. A sample of their vertical acceleration after application of a smoothing component can be seen in Fig. 5.12.

By comparing the bandpass filter signal in Fig. 5.12 (Top) with no denoising applied, and the signal acquired in Huan et al. (2019) (Fig. 5.13) which was used for gait analysis, it can be seen that the bandpass signal is comparable and 'clean' enough to be used for Human Activity Recognition. The cleanliness of the ActivPAL's accelerometer signal is assisted by

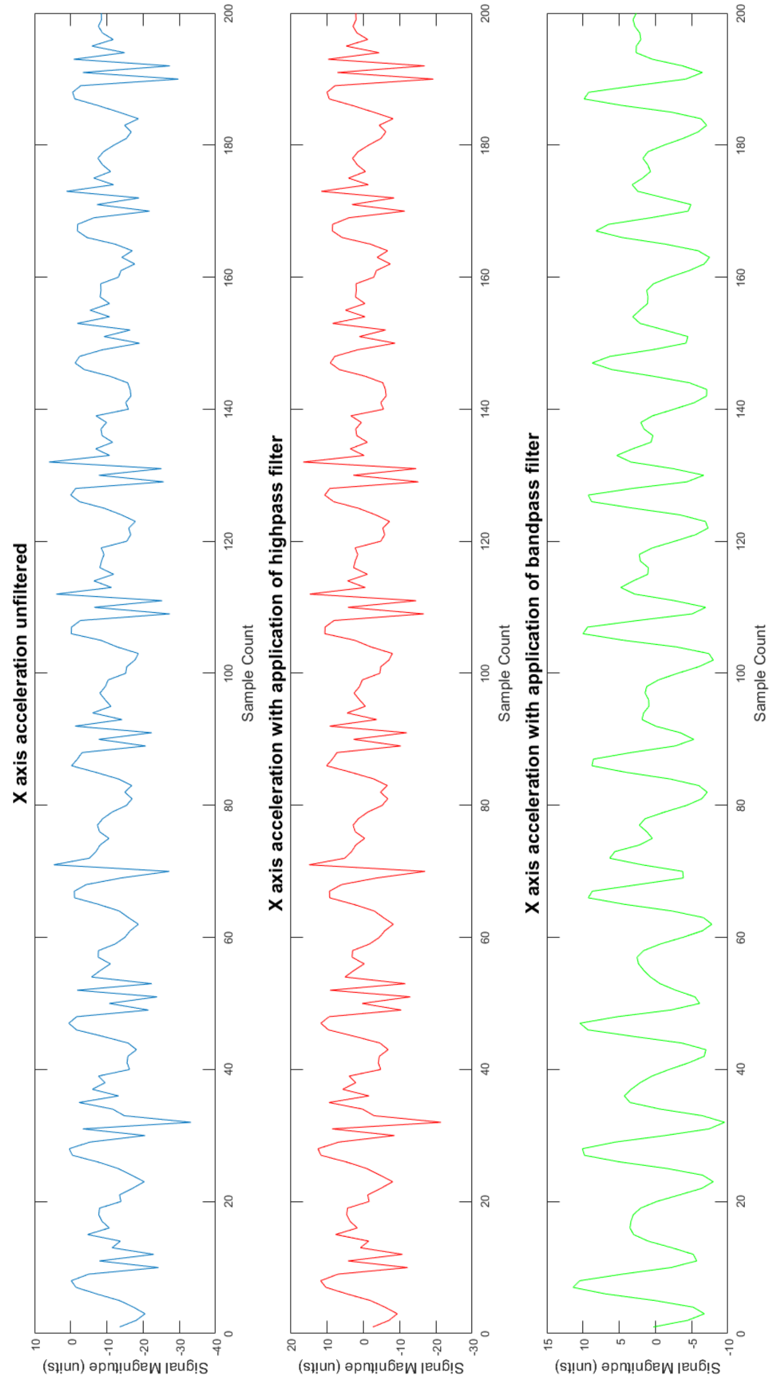


Figure 5.11 (Caption on next page)

Figure 5.11 Unfiltered (Top), Highpass (Middle) and Bandpass (Bottom) readings for a 10-second sample of X-axis accelerometer data from one participant.

its comparatively low sampling frequency, which helps mitigate high frequency artifacts.

5.6.3 Window Segmentation

Another important component in pre-processing is to separate data into smaller chunks of data known as 'windows', where each window can be represented by an identifiable activity. This process is known as data or window segmentation. The three most common approaches to data segmentation are Activity-defined windows, Event-defined windows and Sliding windows (Banos et al., 2014). In Activity defined windows, windows are defined by a change in activity. Hong et al. (2010) triggered windows in triaxial accelerometers when an RFID signal was detected. For instance, if the user picked up a toothbrush, an RFID signal from the toothbrush would be detected by a glove-worn RFID reader, triggering a sampling window. Sensor readings from body-worn triaxial accelerometers would then be used to determine whether the user actually brushed their teeth or simply picked up the brush. Event-based windows are conceptually similar, as windows are defined by the detection of an event. In Zijlstra (2004) for example, external observers marked the start and end of an activity by pushing a button which triggered a signal in a portable measurement system which concurrently read the accelerometer data. The most common segmentation approach is the Sliding window (Banos et al., 2014). In this approach, the data is segmented into windows with fixed sample sizes. The windows can overlap with each other such that some data samples are shared between two windows (Gjoreski and Gams, 2011; Nam and Park, 2013). Dehghani et al. (2019) reviewed studies that included both approaches and found neither segmentation technique was more favourable than the other; overlapping was only found to be the superior technique when the cross-validation of the machine learner was subject-dependent. The number of samples in a window define the window length. In HAR, this window length

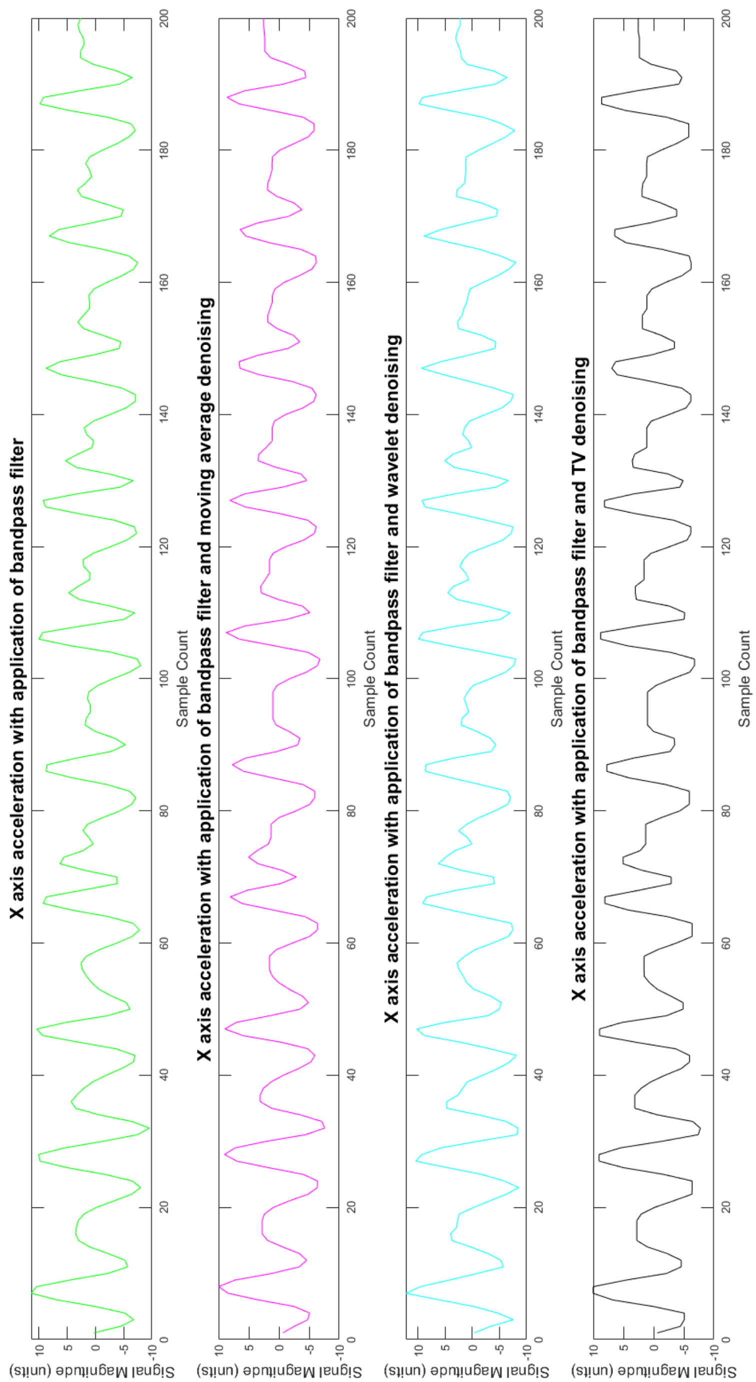


Figure 5.12 (Caption on next page)

Figure 5.12 Excerpt of X axis accelerometer signal after bandpass filtering over 10 second period for one participant. From Top to Bottom: Bandpass signal | Moving average Filter with 2 samples averaging width | Wavelet Denoising of Accelerometer signal. Parameters: Symlet 4, Minimax denoising, Hard Threshold | TV Denoising of Accelerometer signal with lambda = 0.9 and Number of Iterations = 50

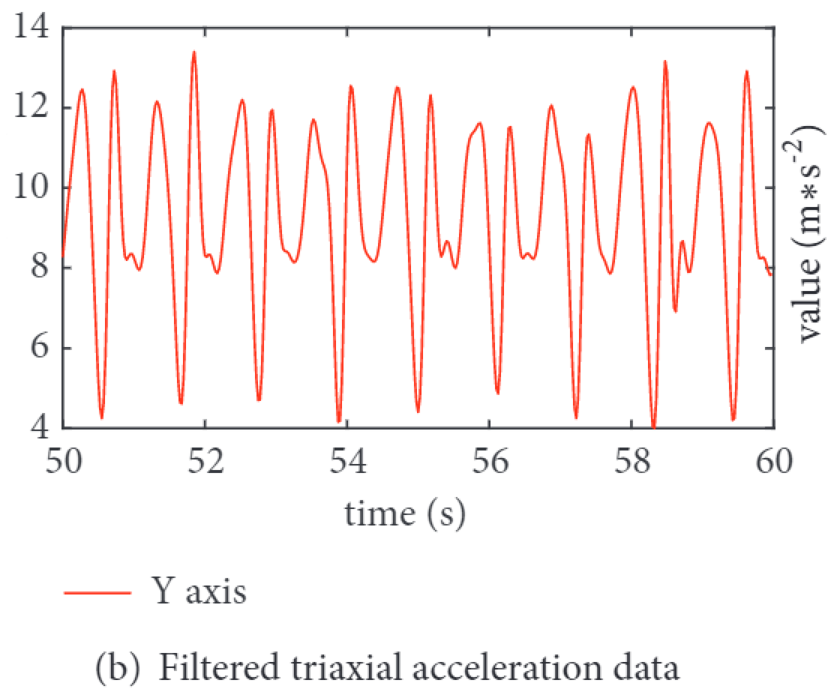


Figure 5.13 Sample of vertical acceleration of thigh-mounted smartphone from Huan et al. (2019). Copyright © 2019 Zhan Huan et al. The original source is licensed with C.C 4.0 granting reuse of images.

should typically range from 1 to 10s (Preece et al., 2009), although it can be difficult to determine an optimal length prior to obtaining results. If the window length is too short, there is the aforementioned risk of splicing activities carried out over longer periods of time, and if it is too long, then it risks containing an overlap of different activities.

In this investigation, activity-defined windows were briefly trialled as a potential approach. The ActivPAL software PAL Analysis contains the option to export "Event Data", in which the cumulative step count of the individual is tracked over time via the VANE algorithm's analysis of the accelerometer data. By utilising the timestamp at which these steps occurred, there was potential to sequester data as a function of steps. After some brief experimentation, it was found that the VANE algorithm was not reliable when it came to counting steps that were traversed on stairs; in the event data, there would occasionally be no steps counted during the stair movement, and the subject would appear to be standing still for the entire period. There is further investigation of VANE's stepping count in chapter 8.2. Instead, it was opted to use sliding windows. In regard to the window length, a sampling window of 2 seconds (40 samples) with no overlap was chosen. A two second window length would be adequate to capture at least one step, and the choice of no overlap was due to being non-consequential compared to other design choices in the machine learning process (Dehghani et al., 2019). Longer sampling windows not only decreased the number of samples available for training, but also caused one activity label ("gravel, uphill") to be removed entirely.

Each individual sample of accelerometer data contained an associated annotation label that was determined by timestamps generated by the VoTT software. As a consequence of segmentation, within a single segment window it was possible to have more than one activity. To deal with this scenario, a function was written such that within each segment window, all activities are identified and tallied. If the largest class constituted 80% or greater of the total samples within the window (Kwon, Abowd, and Plötz, 2019), that window would be annotated with the majority class; if the largest class constituted less than 80% on the other hand, the segment would be discarded from analysis. Further, any segmentation samples that contained even a single instance of a "null" class were automatically discarded from analysis.

5.7 Building towards a Feature Extraction Matrix

Features were conceived from two processes: the first process was to use the relevant features that were discussed in Chapter 4.2.4. The second process was an exploratory data analysis of the accelerometer's time series signals, which takes inspiration from Khan, Siddiqi, and Lee

(2013)'s HAR paper. Since each segment had its corresponding annotation, it was possible to directly acquire the accelerometer signals corresponding to each activity, visually analyse the differences in the signal, and then translate those differences into features that could capture these differences in an objective, numerical form. Naturally, with over 25,000 samples of raw data, it was not feasible to look at every single sample and try and draw inferences. There was a brief attempt to amalgamate all data for a particular activity class into a single figure using an overlay plot. To try to normalize the signal relative to the gait cycle, the window length was dynamically varied using ActivPAL's VANE algorithm, using the data points corresponding to the start and end of a step as the limits of the window. Unfortunately, as Fig. 5.14 demonstrates, this experimentation proved the method as unsuccessful; there were unacceptable levels of temporal variance in the acceleration values to try and appropriate all acceleration behaviour in one class into one figure.

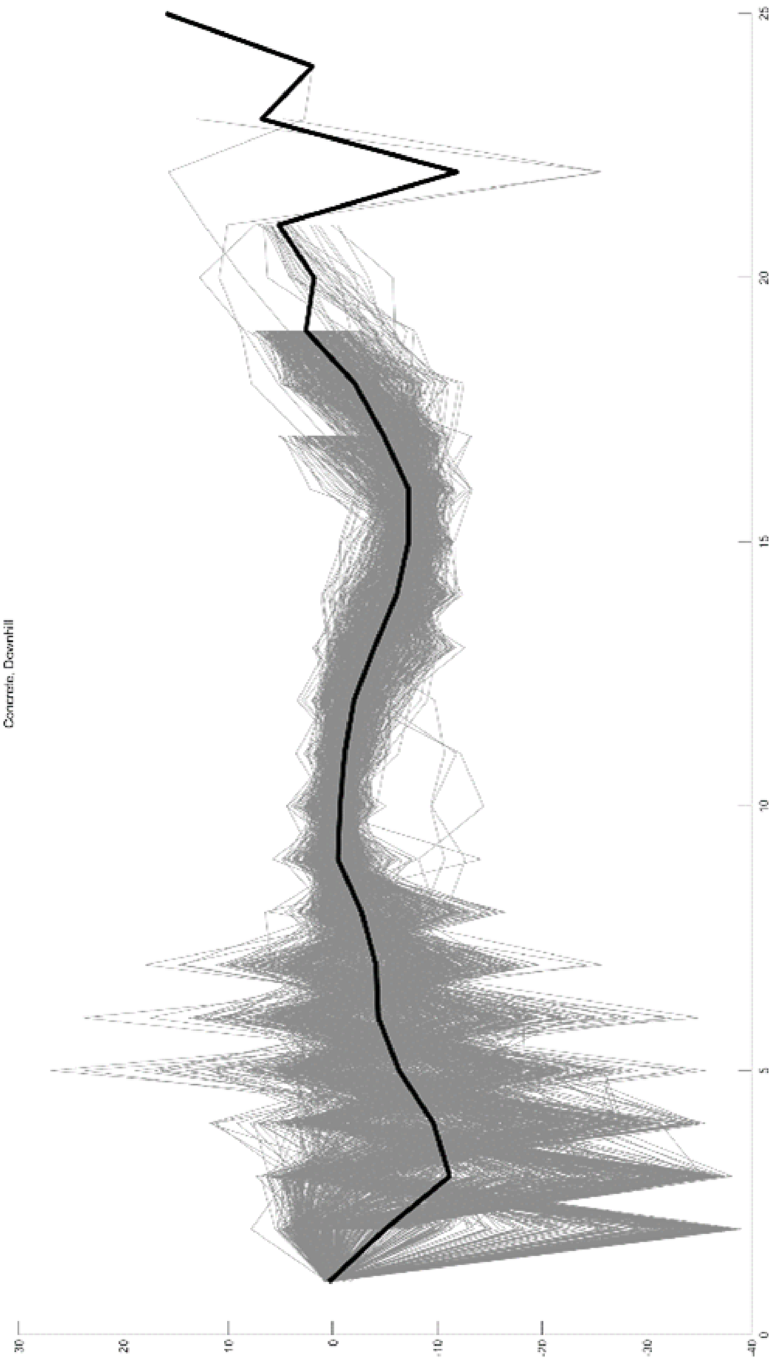


Figure 5.14 (Caption on next page)

Figure 5.14 Overlay plot of the X axis of acceleration for all occurrences of the activity "Concrete, Downhill" for subject H3. The vertical axis represents raw accelerometer values, while the horizontal axis represents the sample window index. The bold plot indicates the average value relative to the sample window index.

The compromise was to acquire five randomly selected samples from each activity class. For each sample, a figure consisting of three sub-figures was generated, with each sub-figure corresponding to each axis of the accelerometer. They were then stored in a notetaking software (Microsoft OneNote™, Microsoft, WA, USA), which was beneficial for creating collages of images with no restrictions on space. Through this collage, it was possible to visually analyse intra-activity variability, and perform pairwise comparisons of the different activity classes. Finally, the scope of data collected was limited viewing data from the amputee subjects, as the "priority" of the machine learning system was to model features based on differences in movement for the amputee population.

5.7.1 Features acquired through Literature

Starting with the features mentioned in Labarrière et al. (2020) as well as the features discussed in Chapter 4.2.5.1, statistical features like the **mean, median, variance, root mean square, norm, mean absolute deviation** and **jerk** (the differential of the accelerometer signal) appear with abundant frequency. As there have been no investigations to dispel the effectiveness of these features within a HAR context, there is no compelling reason to not include them.

From Rosati, Balestra, and Knaflitz (2018), the use of "integral features" relating to the forward movement of gait were adopted. These features are comprised of: **"Mean of integral of Superior-Inferior (SI) acceleration", "Mean of double integral of SI acceleration", "Mean of integral of Anterior-Posterior (AP) acceleration", "Root Mean Square of integral of AP acceleration", "Root Mean Square of double integral of AP acceleration"**. Using the axes of the ActivPAL during gait, SI and AP directions can be approximated as the X and Z axis respectively. As integrations of acceleration, these features relate to the participant's velocity and displacement. Since the integration process introduces an unavoidable integration error (which is magnified under double integration), the resulting values cannot be used to dictate the subject's exact velocity or displacement. Rather, these features provide a rough estimation of velocity and displacement which should be particularly noticeable during stair movements when there will be a change in walking speed and walking direction.

In a similar vein of thought to the inclusion of integral features, the **"eigenvalues of dominant direction"**, as described in Zhang and Sawchuk (2011) were included in the feature matrix, applying to both the acceleration and the jerk. These features are acquired by extracting a covariance of the input signal in a triaxial matrix format, then decomposing into a 3x1 eigenvector. The "dominant" eigenvalues are the second and third values from this eigenvector. The theory of applying these features is that, as a person ascends or descends a hill or flight of stairs, these values should change as a result of the subject altering their vertical trajectory.

The final set of features acquired through literature are the **Mel Frequency Cepstral Coefficients** (MFCCs). Using San-Segundo et al. (2016) and various MFCC tutorials for guidance on parameters, the Mel-spectrogram is formed from an inverse Short-Term Fourier Transform of the input signal. The resulting spectrogram is given absolute values to avoid introducing complex numbers, after which the first 13 cepstral coefficients are extracted. This applied to each axis separately, and each coefficient acts as its own feature, resulting in 39 new features total.

5.7.2 Features acquired through Exploratory Data Analysis

The first set of features acquired through the exploratory process are derived by inspecting the figures of amputee subject A4's typical acceleration behaviour on flat, uphill and downhill terrain. As concrete is by far the most commonly traverse type of terrain and the only one universally traversed across all participants, inspections were initially made using concrete terrain as a common denominator. The triaxial plots for each type of movement are detailed in Figs. 5.15-5.17.

By inspecting the magnitude and peak values of accelerations, particularly in the X and Z axis, some notable differences can be observed. For instance, in downhill movement (Fig. 5.17), the troughs in the X axis typically have larger negative values than in flat and uphill movement. As the subject's foot must travel further to reach the ground in downhill movement, this results in a greater applied force by the foot, and by using Newton's Third and Second Laws, this dictates a greater deceleration as experienced by the ActivPAL. To avoid tripping during downhill movement, the subject must also have greater control over their forward leg thrusts in order to maintain a steady walking pace, which results in the smaller magnitude of accelerations in the Z axis. In contrast, uphill movement (Fig. 5.16) requires similar biomechanical demand to an increase in walking speed; the trailing ankle requires more push-off force to clear the raised surface, and the leading leg requires greater rotational moments in the knee and hip extensor resulting in greater values of peaks and troughs in the Z axis and troughs in the X axis. These implications are supported by Orendurff (2016), who further note that these effects will be especially present in subjects with lower limb amputation. Thus, features relating to the peaks and magnitudes of movement are included. These features are the: **Maximum, Minimum, Range, Energy, Entropy, Power, Quartile values, Interquartile Range, Crest Factor** and **Signal Magnitude Area.**

Additional features were devised from comparing concrete, flat movement to upstairs and downstairs movement in amputee subjects, again using subject A4 for illustration; their plots are shown in Figs. 5.18-5.19.

During stair movement, it can be expected that there will be significant changes in hip and knee moments, ankle flexion angles, and ankle dorsiflexion angles comparing upstairs to downstairs and ground movement (Protopapadaki et al., 2007; Riener, Rabuffetti, and

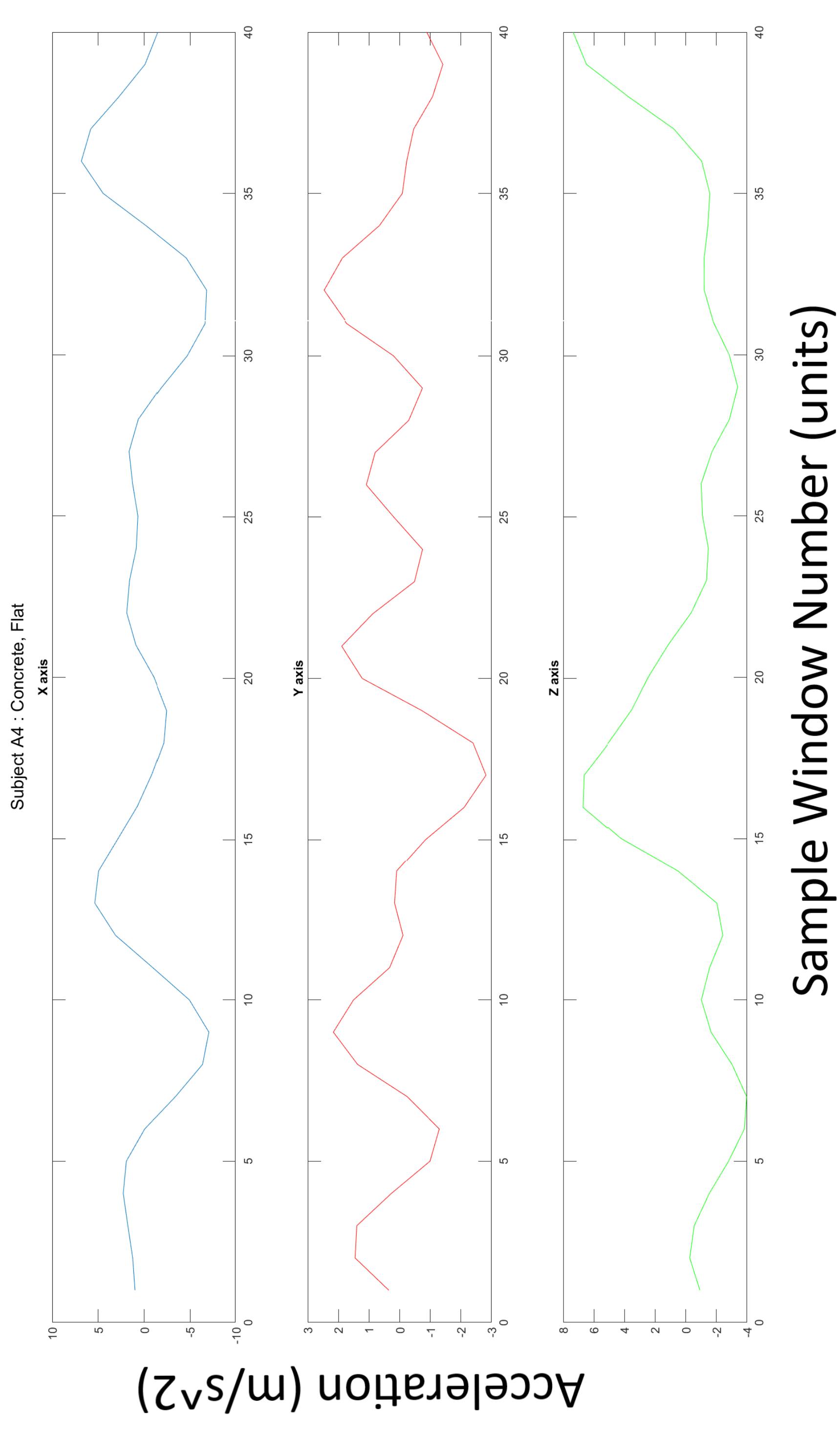


Figure 5.15 Sample of Triaxial data of "Concrete, Flat" activity for participant A4

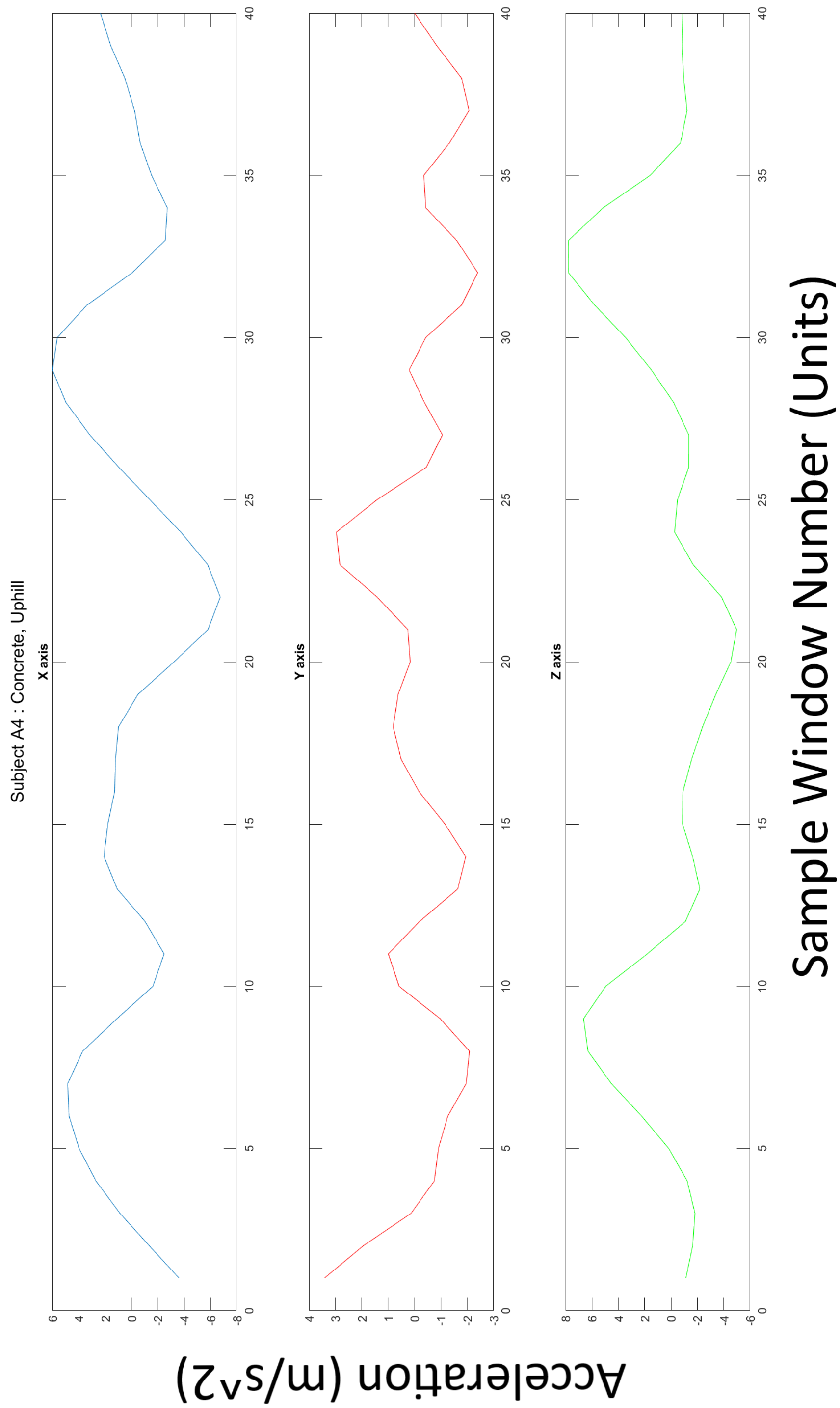


Figure 5.16 Sample of Triaxial data of "Concrete, Uphill" activity for participant A4

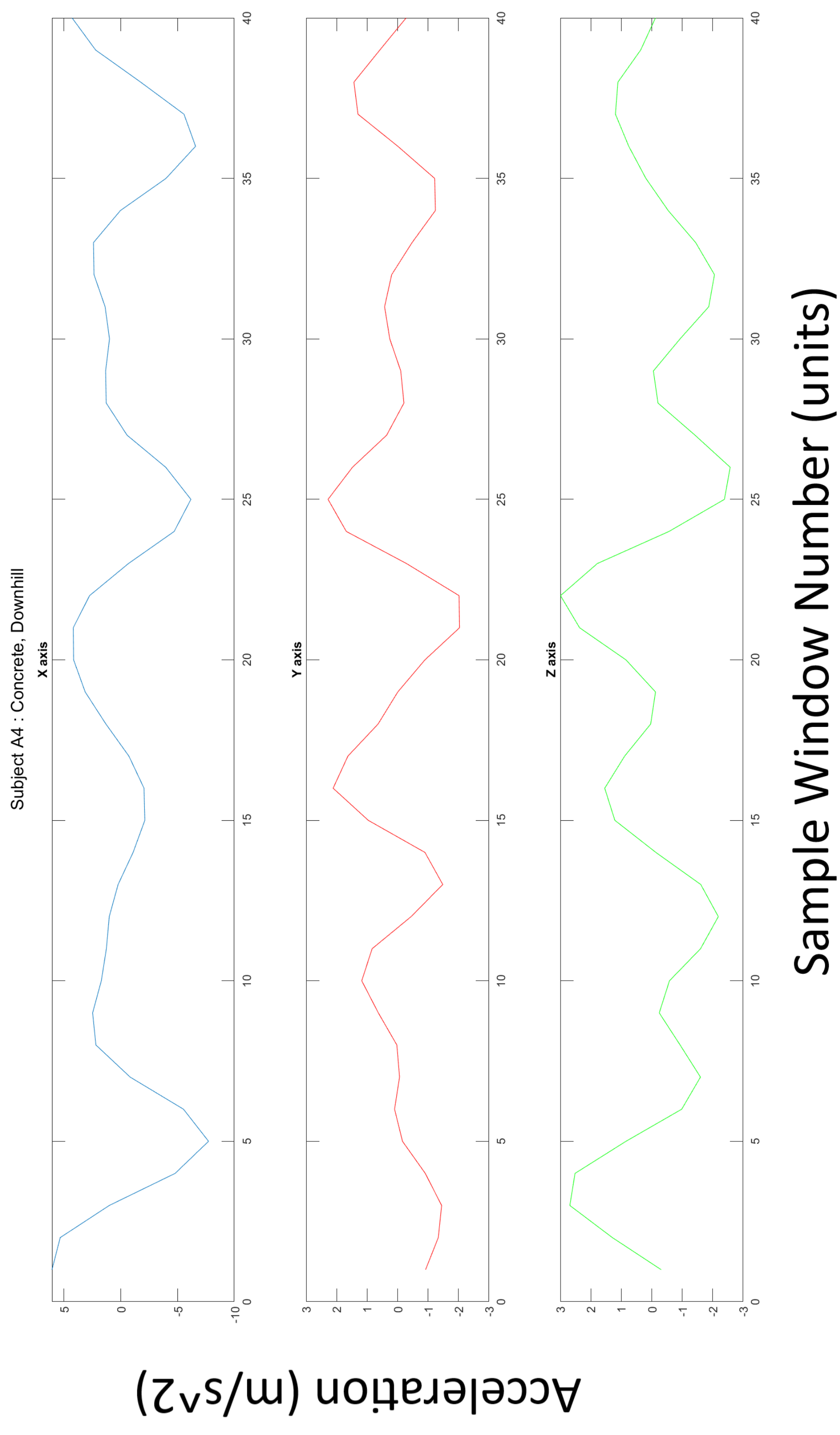


Figure 5.17 Sample of Triaxial data of "Concrete, Downhill" activity for participant A4

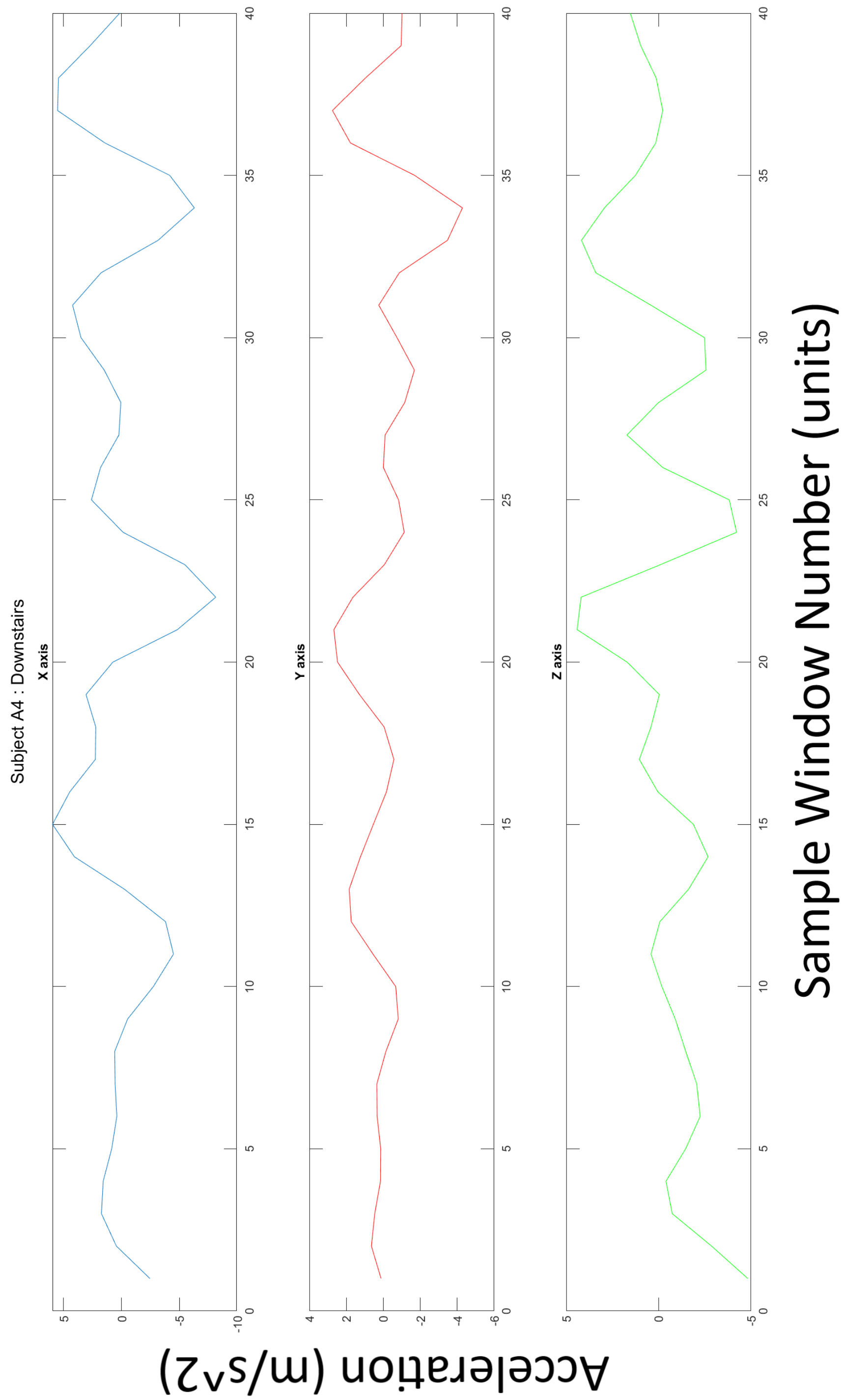


Figure 5.18 Sample of Triaxial data of "Downstairs" activity for participant A4

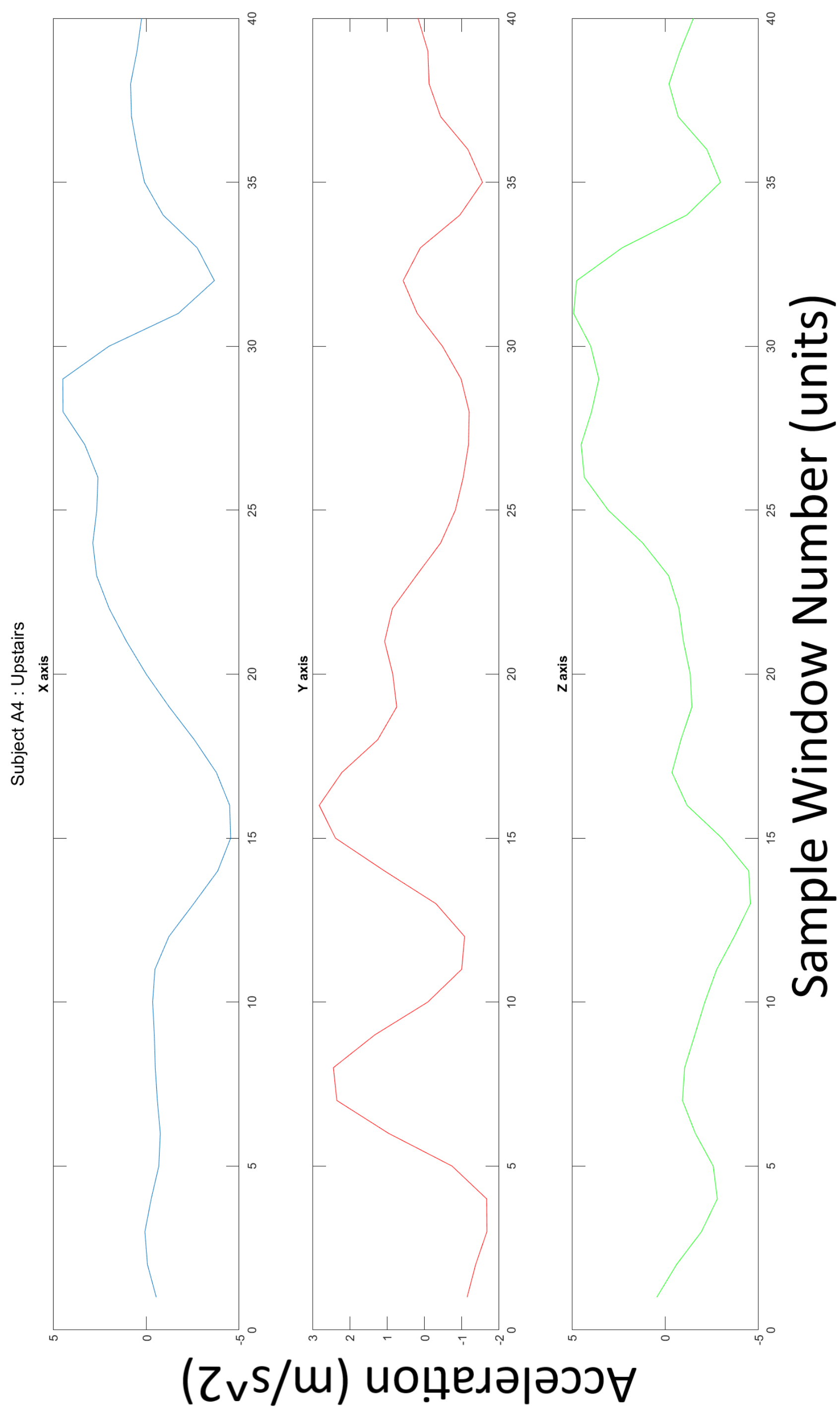


Figure 5.19 Sample of Triaxial data of "Upstairs" activity for participant A4

Frigo, 2002). These findings, which are based on non-amputated population data, should extend to ILLAs, especially in the unilateral amputees where there will be inherent asymmetry between sound and prosthetic legs (Clark, 2018; Powers et al., 1997; Powers, Rao, and Perry, 1998). These findings implicate changes in acceleration regularity, as well as a change in relationship primarily between the X and Z axes of the ActivPAL. These findings can therefore be quantified using **Kurtosis** and **Interaxis Correlation Coefficients**. Interaxis relationships with the Y axis were also included based on a thought that during stair movement, there may be some subtle lateral movement in the leg to help raise the foot over a step during ascent (Harper, Wilken, and Neptune, 2017). Moreover, during stair movement, there will be a natural decrease in walking speed as the subject must adjust their gait pattern to traverse the stairs in a safe manner. As a result, the sampling window will tend to capture only one or one-and-a-half steps compared to two in ordinary gait. This leads to a change in temporal regularity in the signal, which can be quantified through **Skewness**. Because of the expected change in walking frequency, these changes can be reflected in the frequency spectrum via the application of **spectral energy, spectral entropy,** and **spectral centroids**.

The final set of features were devised from analysis of level ground walking on concrete and grass, with the intention of finding features suitable for distinguishing hard and soft terrains. An exemplary timeseries plot of the accelerometer signal under grassy terrain for subject A4 is included in Fig. 5.20.

Distinguishing activities on different terrains when the inclination is equal proved to be a much more challenging task. Unless the terrain is actively impeding movement (for example through uneven ground (Gates et al., 2012; Voloshina et al., 2013)), changes in biomechanics will be minimal. The author posits that the most consistently identifiable change in gait on softer terrain is a reduction in the Ground Reaction Force (GRF), due to the ground absorbing more of the impact from the foot at initial contact, and subsequently requiring slightly more power in the ankle to propel the foot during toe off. A change in the GRF

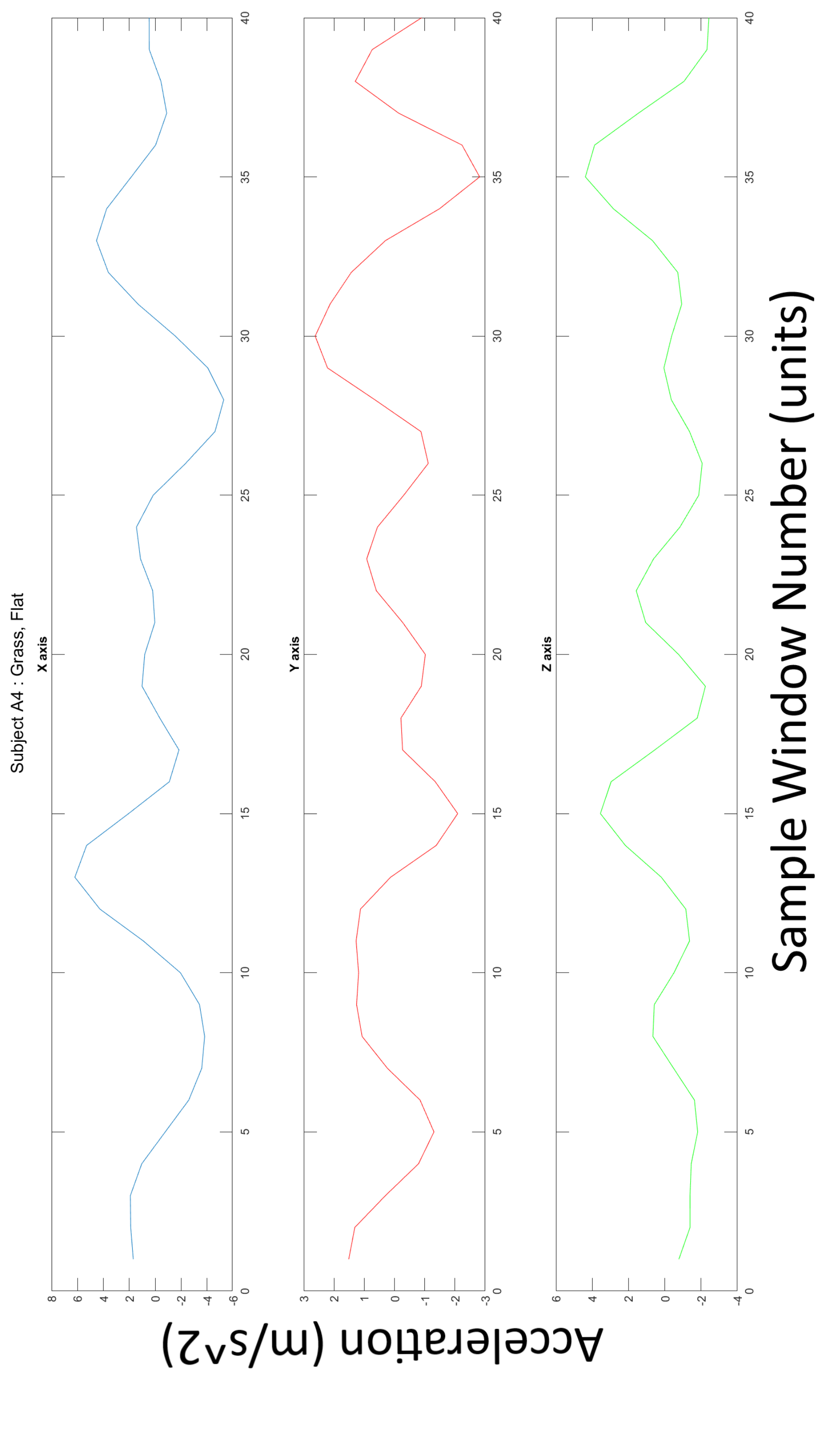


Figure 5.20 Sample of Triaxial accelerometer data of "Grass, Flat" activity for participant A4

should lead to subtle alterations in the oscillatory accelerometer behaviour during these events, which can be best analysed in a domain that considers both time and frequency: the wavelet domain. Unfortunately, there is currently no papers which have investigated wavelet features in the context of distinguishing different terrains during gait, and so the closest approximation was to utilize the wavelet features described in Preece et al. (2009). When it came to deciding which set of wavelet features was most suitable for the investigation, it was decided to include all sets of features. None of the wavelet feature sets used in Preece et al. (2009) had a standout performance, and it was desirable to capture as much detail as possible regarding the subtle changes in terrain variation via the wavelet domain. Furthermore, it was undesirable to have experimentation of the wavelet features inclusion or exclusion become a distraction from the focus of the thesis. Thus, all 76 **wavelet coefficients** as described in table 1 of Preece et al. (2009) were adopted for use. The reader is encouraged to refer to Preece et al. (2009) for an explanation of each set of wavelet features.

Finally, no new features were conceptualized when comparing one hard terrain to another hard terrain, and likewise one soft terrain to another soft terrain. There were no consistently identifiable metrics that could translate differences between the terrains, and the intra-terrain variety created additional caveats. It was simply hoped that the introduction of the aforementioned features would be enough to try and form some level of distinguishment through some complex feature interactions not perceptible by human interpretation. Examples of stone and sandy terrain are included in Figs. 5.21-5.22 for reference.

5.8 Feature Acquisition Results

A condensed list of statistical, frequency and wavelet domain features calculated in each segment of the accelerometer data and their associated formulas are detailed in Table 5.4. A total of 243 features are calculated. While the dimensionality of the dataset may at first appear alarmingly high, the dimensionality is comparable to some studies like Rosati,

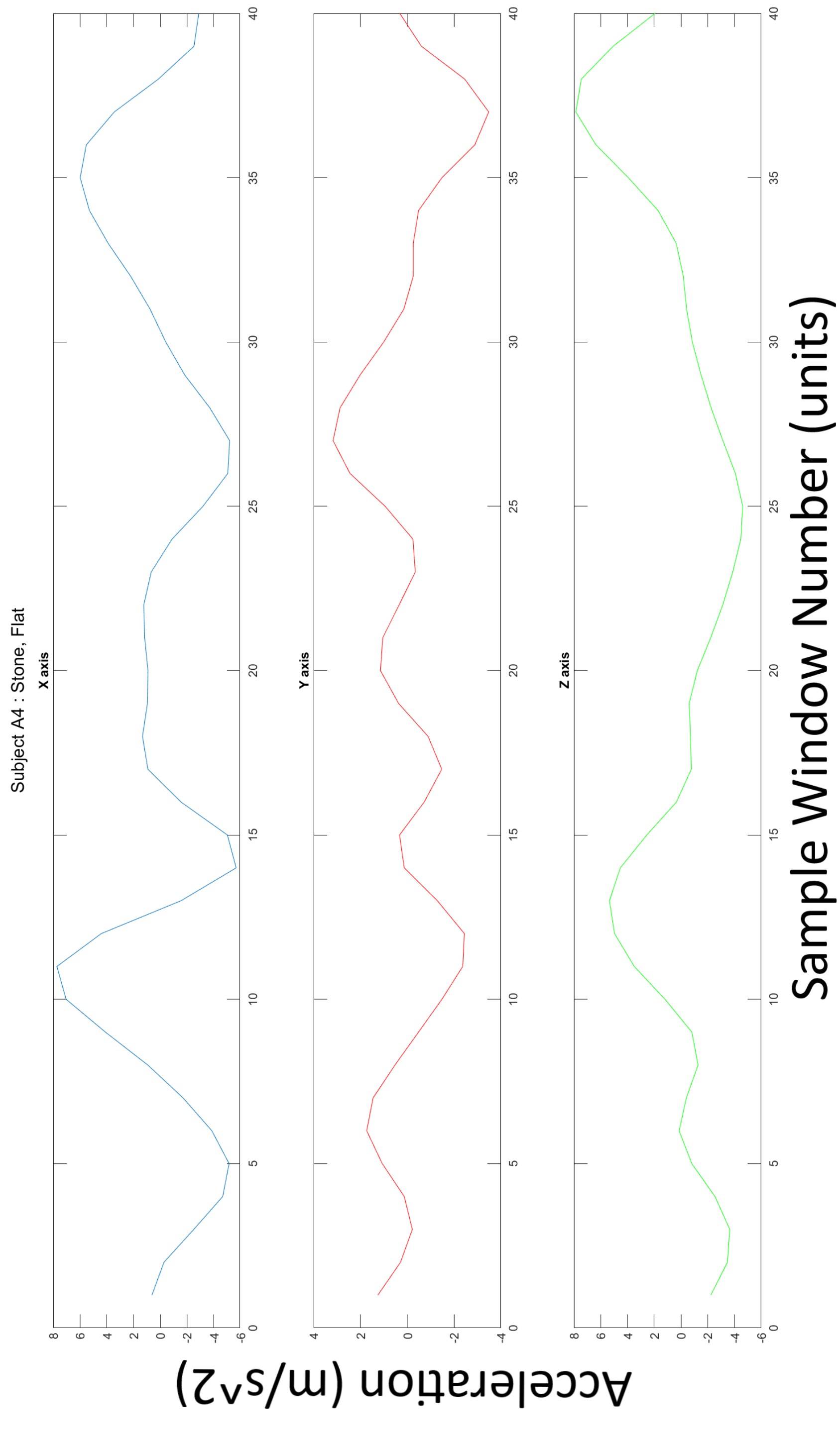


Figure 5.21 Sample of Triaxial data of "Stone, Flat" activity for participant A4

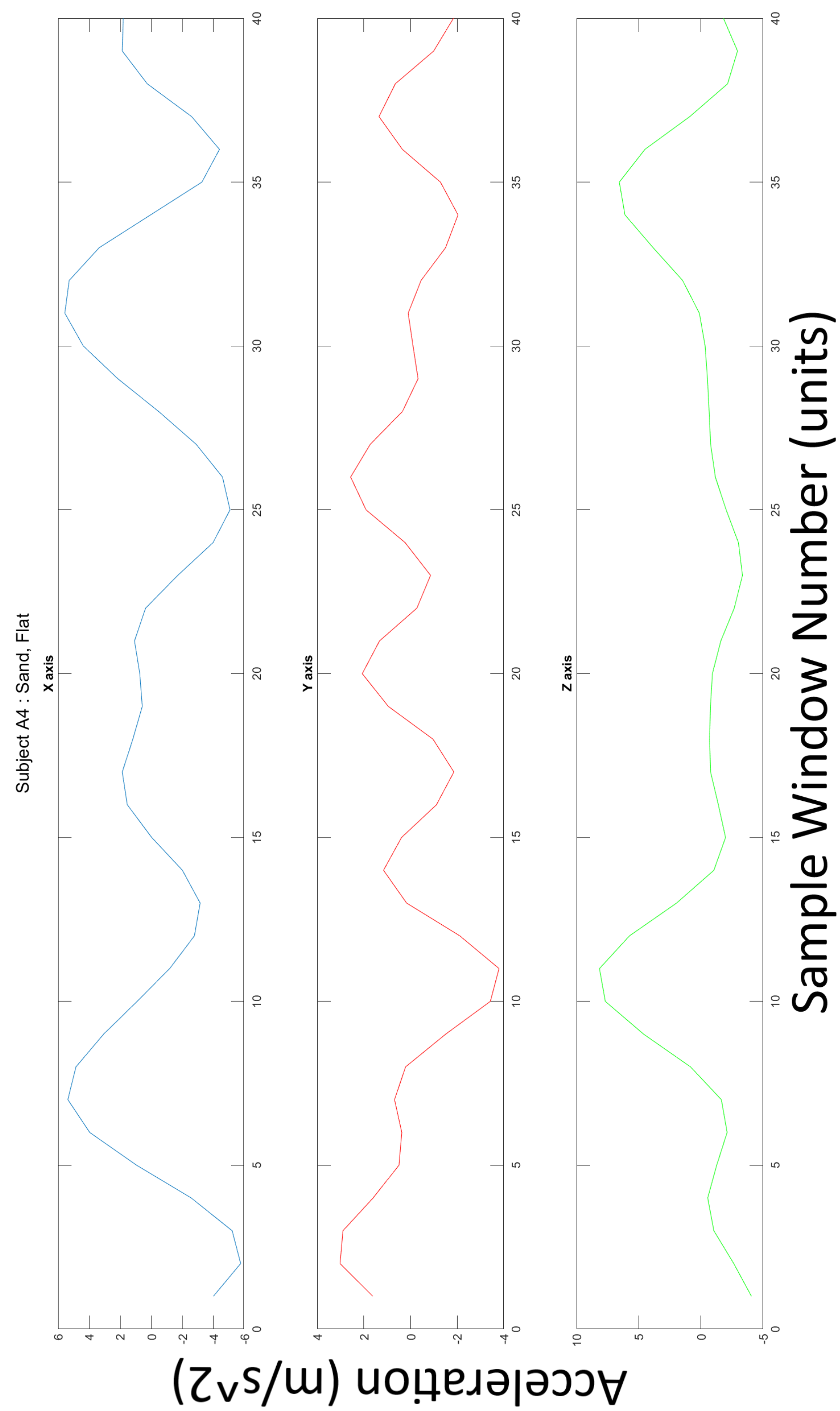


Figure 5.22 Sample of Triaxial data of "Sand, Flat" activity for participant A4

Balestra, and Knaflitz (2018) who were able to still achieve 98% average recognition accuracy in their HAR dataset with around 220 features. In most listed features, there are unique instances of the feature in each axis. For instance, the mean value of acceleration in the X axis is a separate feature from the mean value of acceleration in the Y axis, but both are listed in Table 5.4 as simply "mean". Further, a dozen of the statistical feature calculation processes utilise the differential of the signal (i.e the "Jerk") and the magnitude of the signal as additional features. In certain cases, the definitions of the features are too complex to include within the table, and instead are pointed to the literature for more information. A full list of the features can be found in the author's PURE repository for this thesis (Jamieson, 2021).

Table 5.4 List of machine learning features used in the investigation.

Feature	Explanation
Statistical & Time Domain Features	
Mean	$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$
Median	$Med(X) = \frac{\left(X\left[\frac{n-1}{2}\right]+X\left[\frac{n+1}{2}\right]\right)}{2}$
Variance	$\sigma^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$
Root Mean Square	$RMS(X) = \sqrt{\frac{1}{n}(\bar{X})^2}$
Crest Factor	$c(X) = \frac{\lvert X_{peak}\rvert}{RMS(X)}$
L1 Norm	$L1(X) = \sum_{i=1}^{n} X$
L2 Norm	$L2(X) = \sqrt{\sum_{i=1}^{n} \lvert X\rvert^2}$
Variance of the sample-wise Norm[a]	$\sigma^2_{norm} = \sigma^2(L2(X_i, Y_i, Z_i))\ ;\ i = 1, 2 \cdots n$
Skewness	$SV(X) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^3}{(N-1)*\sigma^3}$
Kurtosis	$KV(X) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^4}{(N-1)*\sigma^4}$
25th Quartile	$Q_{25}(X) = X[\frac{1}{4(n+1)}]$
75th Quartile	$Q_{75}(X) = X[\frac{3}{4(n+1)}]$
Interquartile Range	$IQR(X) = Q_{75}(X) - Q_{25}(X)$
Maximum	$Max(X)$
Minimum	$Min(X)$
Range	$Max(X) - Min(X)$
Mean Absolute Deviation	$MAD = \frac{1}{n}\sum_{i=1}^{n} \left\lvert X_i - \bar{X}\right\rvert$
Signal Magnitude Area	$SMA = \frac{1}{n}(\sum_{i=1}^{n}\lvert X_i\rvert + \sum_{i=1}^{n}\lvert Y_i\rvert + \sum_{i=1}^{n}\lvert Z_i\rvert)$
Energy	$E(X) = \sum_{i=1}^{n}\lvert X_i\rvert^2$
Power	$P(X) = \frac{\sqrt{\sum_{i=1}^{n}\lvert X_i\rvert^2}}{n}$
Entropy	$H(X) = \sum_{i=1}^{n}(P_i)\log_2(P_i)\ where\ P_i = \frac{\frac{X_i}{Max(X_i)}}{\sum_{i=1}^{n}(\frac{X_i}{Max(X_i)})}$
Integral Features[b]	Rosati, Balestra, and Knaflitz (2018)

Inter-axis Correlation Coefficients	$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2(Y_i - \bar{Y})^2}}$
Eigenvalues of Dominant Direction	Zhang and Sawchuk (2011)
Frequency Features	
Spectral Energy	E(FFT(X))
Spectral Centroid	$SC = \frac{\sum_{k-1}^{n}\|X(i,k)\|*f(k)}{\sum_{k-1}^{n}\|\|X(i,k)\|}$
Spectral Entropy	H(FFT(X))
Cepstral Coefficients	San-Segundo et al. (2016)
Wavelet Features[c]	
Tamura Coefficients	Tamura et al. (1997)
Nyan Coefficients	Nyan et al. (2006)
Sekine Coefficients	Sekine et al. (2000)
Wang Coefficients	Wang et al. (2007)
Fractal Dimension	Sekine et al. (2002)
Preece Coefficients	Preece et al. (2009)

Abbreviations: X – one dimensional time series input signal of acceleration. || n – number of samples within a segment || σ – standard deviation of time series signal || |Xpeak| - the magnitude of an identified peak within a time series signal || Y – a secondary time series input signal of acceleration that is not the same as the primary input signal (X) || Z - a tertiary time series input signal of acceleration that is not the same as the primary input signal (X) or the secondary input signal (Y) || $\int_0^T |X|\,dt$ – trapezoidal integration of time series signal over sampling period T where T is the inverse of the sampling frequency || FFT(X) – Fast Fourier Transformation of signal X || k – frequency bin from a FFT || f(k) – frequency of bin k in Hz

[a] - to elaborate, the variance of the sample-wise norm is calculated by acquiring the L2 norm of the X,Y and Z axis for each sample of the input signal from 1 to n. σ^2_{norm} is the variance of the calculated norms across the entire segment.

[b] - The integral features used from Rosati, Balestra, and Knaflitz (2018) are: "Mean of integral of Superior-Inferior (SI) acceleration", "Mean of double integral of SI acceleration", "Mean of integral of Anterio-Posterior (AP) acceleration", "Root Mean Square of integral of AP acceleration", "Root Mean Square of double integral of AP acceleration". Using the axes of the ActivPAL during gait, SI and AP directions are approximated as the X and Z axis respectively.

[c] - All Wavelet Features are utilised from Table 1 of Preece et al. (2009) and is the recommended starting point for understanding the wavelet features used.

The time that is required to extract this list of features bears mentioning. The features are extracted in a Matlab script after filter and segmentation pre-processing. The computation specs for the feature extraction process were previously established in Section 5.3. Fig. 5.23 demonstrates the processing time required to extract all features relative to the size of the dataset, which is controlled by the number of seconds contained within a single window segment. The processing time was not ideal with the chosen window sample size of two seconds, requiring around 15 minutes to perform calculations. Boosting the sample size window to 10 seconds still required 3 minutes to perform the extraction process, as illustrated in Fig. 5.23. On the other hand, as the system did not require Realtime processing, the slow processing speeds was not of vital concern.

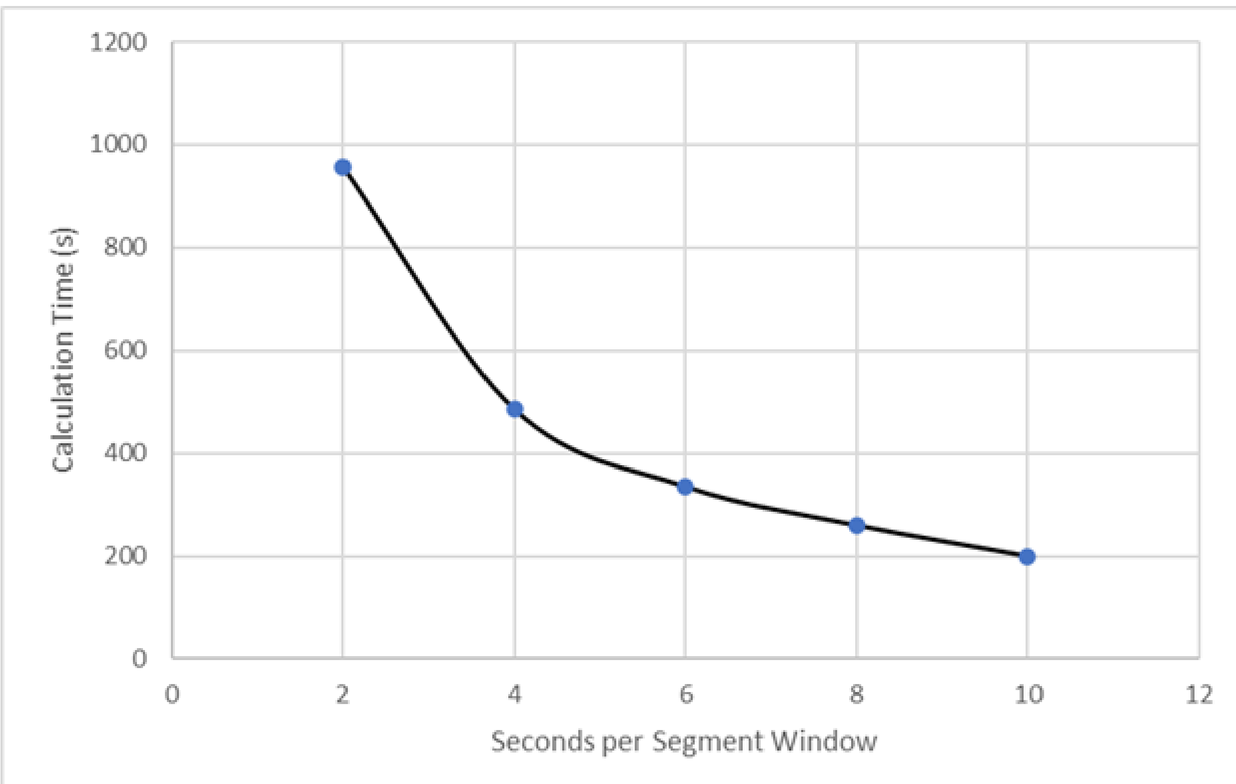


Figure 5.23 Processing time for feature extraction relative to the length of the sampling window

To understand which features are the most processing intensive, Matlab has additional functionality which allows the user to time the calculation of each process. The functions

Table 5.5 List of the 10 longest feature extraction processes

MATLAB Feature Extraction Function	Time for Calculation (s)
wavedec	335
dwt	197
cepstral_feature_function	192
wfilters	145
prctile	118
stft	102
wavemngr	85
stftParser	74
AutoCorr_Features_function	71
appcoef	69

which required the most processing time are listed in the Table 5.5. The total time for the feature calculation process is 833.5s for 2 second sampling windows, requiring approximately 0.03s of calculation time per sample window. The most time-consuming feature calculations revolve around features in the wavelet domain (as evidenced by *"wavedec"*, *"dwt"*, *"wfilters"*, *"wavemngr"*, *"appcoef"*) and the frequency domain (*"stft"*, *"stftParser"*, *"cepstral_feature_function"*), though some statistical processes also consume a considerable amount of time (*"prctile"*, *"Autocorr_features_function"*). For a full explanation of what these functions are, refer to Appendix B.

5.8.1 Feature Scaling

It is common practice in machine learning to scale the magnitudes of features. Some classifiers utilize a distance metric in deriving classifier boundaries, and so it is important to ensure that these distances are normalized respective to their magnitude (scikit-learn, n.d.).

There are various accepted methods of feature scaling, and the one chosen in this investigation is standardization. Each value of the feature f, x_f, is replaced by its standardized (Z) value $x_f{}'$ by subtracting the mean μ_f and dividing by the standard deviation σ_f:

$$x_f{}' = \frac{x_f - \mu_f}{\sigma_f} \tag{5.2}$$

This ensures that all features have a zero mean and standard deviation of one. In practice during the subsequent investigations, standardization is applied separately to the training and testing sets of feature data to ensure values from the testing set do not influence the values in the training set.

5.9 Discussion

5.9.1 The transition from "Ideal" Activity Monitoring Outcomes to Practical Activity Monitoring Outcomes

Recalling back to Chapter 3.2, a list of ideal activity monitoring outcomes was devised. Due to the nature of the study, it became unfortunately impractical to measure all of the listed outcomes. The outcome measures that were not practical for analysis within the confines of the dataset were:

1. Indoor/Outdoor movement: As the study was not permitted to film indoors, it would be very difficult to provide ground truth. Other methods of providing validation, such as an activity diary written by the subject, or using GPS information would have privacy and ethical concerns and would very likely be disallowed by the ethics committee. Previously, the author carried out some rudimentary trials of different types of steps in a laboratory environment with non-amputee volunteers, writing down timestamps as activities were started and finished. However, this proved to be a demanding task as timestamps had to be repeatedly written down for every single activity transition, leading to imprecision in the timestamps.

2. Cycling: This would require having to ensure that all participants owned and could ride a bike and would create additional safety concerns about using the bike e.g on public roads.

Noticeably, there are three additional outcome measures which have not been discussed yet: Walking speed, Wear Time, and Cadence. As these are not conventional activities to be "recognized", they are not included in the machine learning process. Cadence is a derivative measurement of the Step Count over a period of time (typically steps/minute); thus, cadence can be assessed by the step count, which is discussed further in Chapter 8.2. While there was scope to calculate walking speed using only the accelerometer data (Baroudi et al., 2020; Byun et al., 2019; Yang and Li, 2012), a lack of a criterion comparison method meant validation of walking speed was impractical. However, walking speed can be inferred in a clinical setting by acquiring the patient's average stride length and multiplying by ½ of the subject's cadence (Dale, 2012; Perry and Burnfield, 2010). Likewise, being unable to film inside the patient's home for ethical considerations made assessing wear time impossible. Instead, in Chapter 8.3 there is an exercise conducted to determine the accuracy of stationary movement and from this exercise proposes a simple unverified algorithm to detect wear time.

5.9.2 Limitations of the Dataset

Throughout this chapter, there have been some references to the limitations of the dataset; namely that it was not allowed to record indoor activities or to record in unsafe weather conditions. There was also some location bias with the volunteers, and in the case of the ILLA group, there is some volunteer bias mandating that the participants in the study would have a somewhat healthy lifestyle, potentially impacting on gait dynamics during terrain movement (Su et al., 2008). Noticeably, all amputees used a Step-over-Step approach when traversing stairs (each step traverses a single stair), while less fit or experienced amputees may adopt a Step-To approach (places both feet on one stair before moving to the next one), which can result in completely different gait patterns (Lura et al., 2017; Narang et al., 1984). All amputee participants were also transtibial, which inhibits the generalizability of the findings to transfemoral amputees who will have different biomechanical responses during stair ambulation (Schmalz, Blumentritt, and Marx, 2007). Moreover, transfemoral amputees who perform step ambulation with Step-over-Step are less frequent compared to Step-To adopters (Hobara et al., 2011).

An unfortunate situation that arose when analysing activities was that there was a disparity in the type of activities collected between the two groups of participants. Fig 5.24 demonstrates word clouds of activities collected by ILLAs and non-amputated individuals. In the case of non-amputated individuals, there was significantly less variety in terrains traversed compared to the ILLA population. This discrepancy arose mainly due to the locations of the non-amputated individuals; most lived in urban areas where access to places like parks or football fields would require travelling long distances. In contrast, the ILLAs were situated outside of the main City of Glasgow where they had convenient and easy access to parks, beaches, and golf courses. Reflecting on the methodology, it was a poor idea to only "encourage" participants to walk on a variety of terrains, instead it should have been a minimal requirement to at least traverse on grassy terrain to try and provide some contrast of softer terrain data.

Finally, there was some issues regarding the synchronization process between the camera data and the GPS timestamps. As previously stated in Section 5.4, the camera used in the recording contained an internal clock which had a drifting factor of approximately 3 seconds per day relative to the time of the GPS. It was eventually realized that the timestamps of the GPS readings could also be unreliable; the GPS timestamps are acquired from the recording smartphone, which in turn automatically acquires the time through its local mobile network. After some experimentation, it appeared that the camera time relative to the GPS time at the start of a recording would be different from the end of recording, even accounting for clock drift. The design methodology did not anticipate the fact that network clock timing

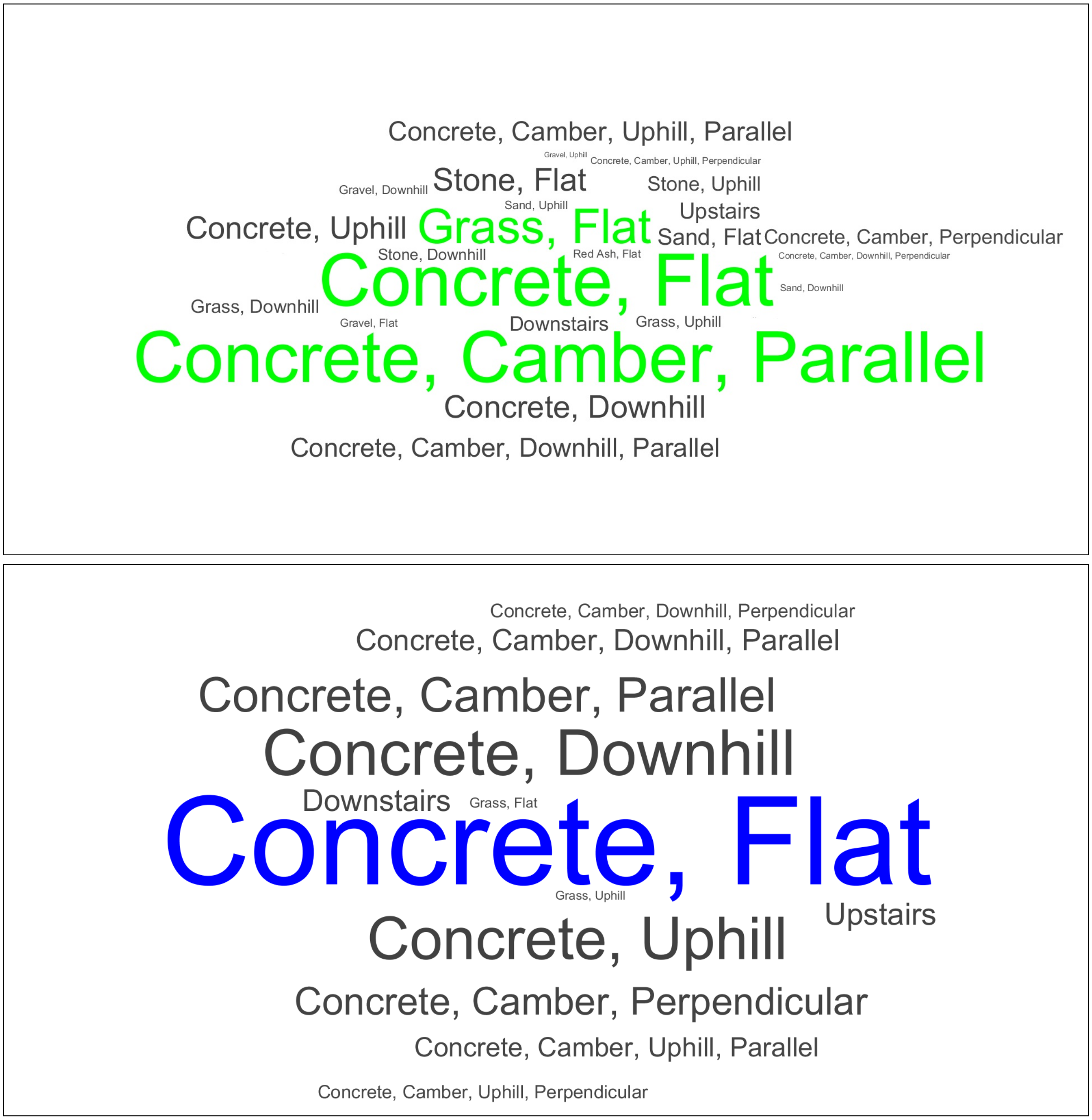


Figure 5.24 Top: Word cloud of activities collected by ILLA participants. Bottom: Word cloud of activities collected by non-amputated participants. Dominant activities are highlighted in green and blue for ILLA and non-amputated participants respectively

is prone to wandering errors and imprecision (Grayson, Shatzkamer, and Wainner, 2009). While this may have led to some mistiming regarding the identification of uphill and downhill data, the error of timing is unlikely to be greater than a handful of seconds, and the “null” angle threshold employed in hill annotation would have further prevented misclassification of uphill, downhill, and flat terrain. In future, this issue should be resolved by turning off the recording smartphone’s option to acquire time automatically via network signals and instead use its internal clock.

5.9.3 Construction of Machine Learning Investigations

Referring to 4.3.7.1, the decision to implement a neural network would depend on the total size of the dataset collected. After collecting all data, the total dataset size was counted using 10 seconds per window segmentation, giving a total of approximately 5,000 samples. This number is comparable to other deep learning studies that have collected their own data, as seen in Table 5.6. Thus, it was decided to try and employ an LSTM network as a comparison point to other supervised classifiers. The LSTM uses the same segmentation and sampling processes as discussed in Section 5.6.3. In Chapter 6, the post-filtered timeseries data from the accelerometer is used as an alternative input method to the LSTM; in those scenarios, a segmented input to the LSTM classifier is 40 consecutive samples of the accelerometer in the X, Y, Z axis or a euclidean vector of the combined axes. The segment is given its label using the same majority rule procedure as before: if a particular activity label populates 85% or more of a 40 sample input, the segment is given that activity label, if not, then the segment is labelled "null" and is discarded from analysis.

In the second half of Chapter 4.3.7, it was discovered that no known investigations have carried out HAR on an amputee population using an unsupervised learning approach. Due to the disparity between the supervised and unsupervised approaches, they were not directly compared to each other, and instead two parallel investigations were conducted. The first

Table 5.6 Table of comparable research that have utilized neural networks and their total data sample sizes

Reference	Neural Network Archetype	Given/Estimated Sample Count for 10 second window
Zebin et al. (2018)	LSTM	2625
Sani et al. (2017)	CNN	4650
Lee, Yoon, and Cho (2017)	CNN	7026
Milenkoski et al. (2018)	LSTM	8250
This Thesis	LSTM	**4940**

investigation compared various supervised learners and an LSTM network to determine which model was the best at predicting activities in ILLAs and non-amputated individuals. The second investigation meanwhile looked at the activities when represented in low dimensional feature space and looks at various cluster algorithms' ability to correctly characterize clusters as activities. Common to both papers, there was also investigations into the appropriate dimensionality reduction techniques. Additionally, there was an analysis of the trade-off between terrain label "resolution" and performance of the systems.

On a final note, it may be noticed that the filter approach was handled in a qualitative analytical manner, rather than an experimental manner. This is because, as Fig. 5.23 illustrates, the feature extraction process consumes a considerable amount of processing time, even if the sampling window increases. As the filtering and segmentation process has to be carried out prior to feature extraction, it is not practical to use these parameters as dynamic variables, whereas any process calculated post-feature extraction (e.g dimensionality reduction or training a classifier) can be experimented with by simply saving the calculated feature matrix in a .csv format and loading in the data when needed. Fullerton, Heller, and Munoz-Organero (2017) compared various filter approaches (Chebyshev vs. Elliptical vs. Unfiltered) across various classifiers and found no significant changes in the average recognition accuracy, implying that the choice of filter will typically have minimized impact.

5.10 Chapter Conclusion

This chapter documented the "backbone" of the HAR process: that is, to explain the processes and challenges involved in the data collection, annotation of the data, handling the sensor information, then applying pre-processing techniques. Finally, the features extracted from the dataset are engineered based on findings from Chapter 4 and exploratory data analysis of some of the collected timeseries samples. The following chapters will detail the "investigative" components of HAR by analysing which features are important to the HAR process, which supervised classifiers can best achieve high recognition accuracy, precision and recall, and which unsupervised approach can best create informative clusters of activity data.

Chapter Six

Supervised Learning of Human Activity Recognition for Non-Amputated Individuals and Individuals with Lower Limb Amputation in Free-Living Conditions

6.1 Introduction

This chapter covers a supervised learning approach to HAR of walking activities carried out by ILLAs and non-amputee individuals using a single thigh-worn accelerometer. The purpose of performing classification of different walking activities was to establish the basis of a reliable activity monitoring system, in which HCPs can track how their ILLA patients remain active over long periods of time and in free-living conditions. There are publicly available datasets of individuals in free-living conditions (Cruciani et al., 2019), but as outlined in Chapter 4.3, there are currently very few studies which have carried out a HAR study in free-living conditions with an ILLA population (Labarrière et al., 2020). Zhang et al. (2019a) and Zhang et al. (2019b) are currently one of the very few studies that are known to have done so; however, their sensor configuration utilised RGB data from a camera as one of the inputs for terrain recognition process, which can have serious ethical implications if applied in a clinical monitoring context (for example, having the camera capture the faces of other people without their consent) (Jung, 2020). Additionally, while locomotion activities are very commonly analysed in ILLA HAR studies, these are typically limited only to walking on level ground, ramps, and stairs (Labarrière et al., 2020). This investigation is, to the author's knowledge at the time of writing, the first to investigate the potential for terrain "resolution" in HAR, at least for an ILLA population (Hashmi et al., 2019; Hu et al., 2018; Russell et al., 2021). The investigations carried out for this chapter are divided into four experiments: the first experiment is a continuation of the investigations carried out in Chapter 5, where it was determined whether feature selection, dimensionality reduction or a combination of both would provide the best method of preparing the feature data for classification. The second experiment took an assortment of classifiers and a deep learning network and used hyperparameter tuning to determine which classifiers were the most suited for recognizing walking activities. Subsequently, the best classifiers from the second experiment were then used in the third experiment to investigate how the terrain "resolution" impacts on classifier performance, and in the fourth experiment, there is an investigation as to whether the data trained on a group of subjects (either ILLAs or non-amputee individuals) could recognize activity carried out by an ILLA.

6.2 Methodology

The methodology for data collection, annotation and pre-processing was already described in detail in Chapter 5. The remainder of this section will describe the methodology around balancing data, configuring the classifiers and how the main experiments were conducted.

6.2.1 Balancing Data

In Chapter 5, activities are given what is referred to as "full terrain resolution", such that each sample is described with its terrain and the condition of the terrain. However, at such a level of detail, this results in 24 activities, thus will adversely impact on the machine learning system due to similarities in features between subsets of similar activities. In experiments conducted in this chapter, the classes are processed with "label consolidation", where similar activities are given the same label. For the first two experiments, the annotated labels were given no terrain resolution, this meant that there were only five classes for potential classification: flat walking, uphill walking, downhill walking, upstairs movement, and downstairs movement. This was designed to give an idea of how the classifiers would perform at the simplest level of classification and is comparable to the labels used in other studies with ILLAs (Labarrière et al., 2020). However, even at the simplest level of terrain resolution, there is a disproportionate balance in activity classes as seen in Fig. 6.1:

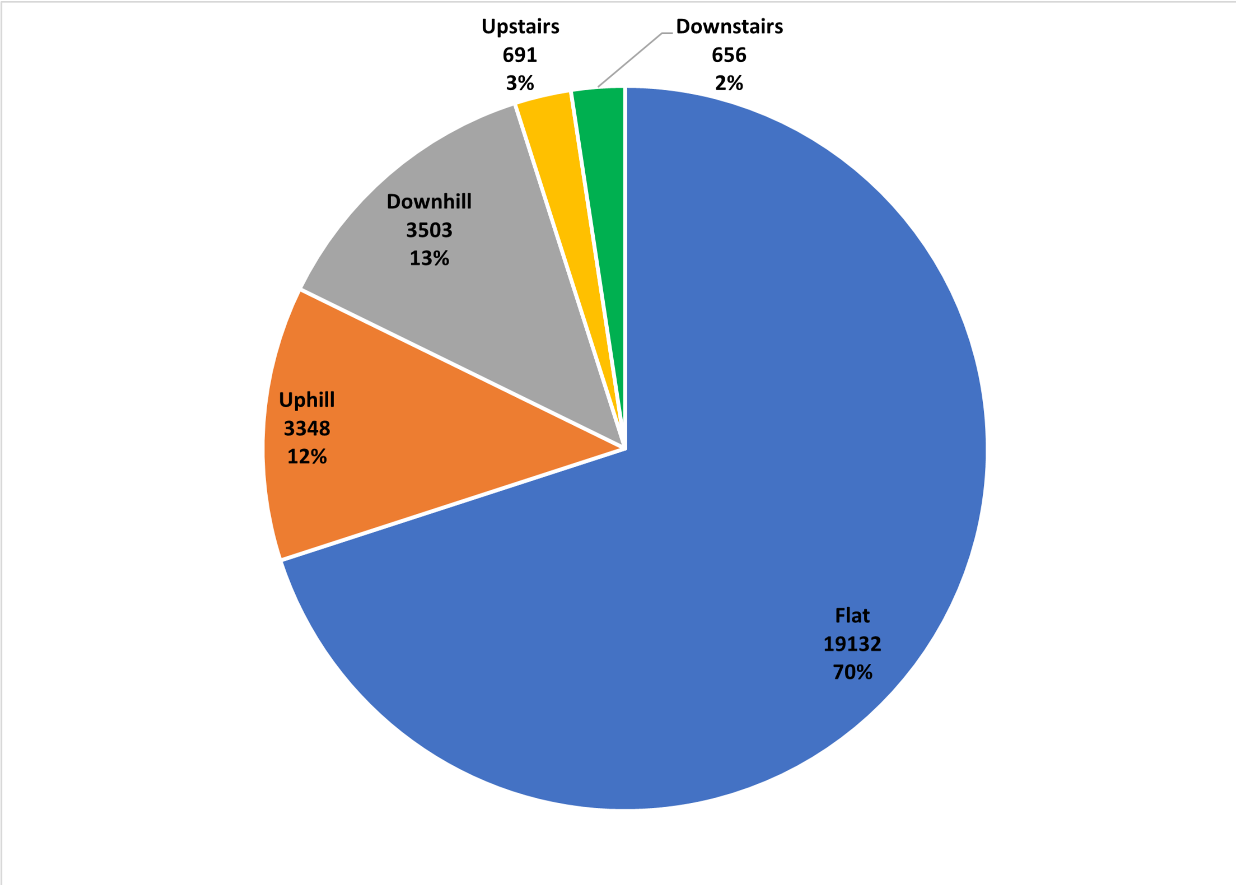


Figure 6.1 Pie chart of frequency and distribution of activities at level 1 of label resolution (5 classes)

In order to avoid having a biased classifier that favours prediction of the majority class (Guo et al., 2008), classes were rebalanced using a sampling technique called Synthetic Minority Oversampling Technique (SMOTE) (Chawla et al., 2002). SMOTE is a variant of oversampling, where feature data attributed to minority classes are synthesized by interpolating the feature data of "nearest neighbour" instances of the same class. The technique is applied to the training partition of the data after the training/validation split of the dataset. If SMOTE is applied prior to the split, the dataset will overfit because the testing partition of the data will contain synthetic samples that are a close approximation of samples relegated to the training partition (Altini, 2015). In other words, the classifier will be testing data that is extremely similar to the data it was trained with, increasing the probability of a successful classification thereby giving an artificial boost in performance. Moreover, SMOTE is not applied separately to balance the testing data because the testing accuracy should be indicative of how the classifier would perform with entirely new data, which should follow similar unbalanced distributions in classes. Regarding the application of dimensionality reduction before or after SMOTE, there is no scientific consensus on the correct order, and it appears that the choice will have minimal impact either way (Blagus and Lusa, 2013). In all experiments of this Chapter, SMOTE was applied prior to any feature extraction or selection processes.

6.2.2 Construction of the LSTM Network

The LSTM model constructed for this investigation was a simple Sequence-to-Label network demonstrated in Fig. 6.2. While there was potential to build a more sophisticated LSTM network with more layers, a simple model helped keep the number of hyperparameters needed for the network to a manageable level and allowed for quick optimization processes in the second experiment. Following guidance from Bengio (2012) and some heuristic experimentation, the most suitable (constant) parameters defined for the network were to perform ADAM Optimization (Kingma and Ba, 2015), set mini batch size to 256, set the initial learn rate to 0.005 with a decrease in the learn rate by a factor of 10 every 2 training epochs, and set the maximum number of epochs to 15.

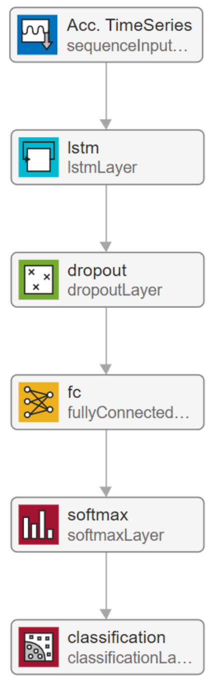


Figure 6.2 Layers of the LSTM Network

In the LSTM classifier, raw training data is also balanced via the application of SMOTE. After completion of the procedures, there was some concern that applying SMOTE to the raw data may have created nonsensical signals, as the SMOTE process will generate new data based on sample index of the timeseries signal. For example, the first element of a SMOTE generated signal for stair movement is generated by random interpolation between all other first element values of the original non-synthetic stair samples. Upon investigation of the raw timeseries signals however, the synthetic signals appeared to have reasonable data points compared to the original samples. Examples of these signals for stair activities are demonstrated in Fig.6.3. In future studies however, the SMOTE process should be replaced with a more appropriate synthetization technique for raw timeseries data. For example, a synthetic signal could be formed by replicating one sample and adding a small percentage of noise relative to the magnitude of the signal to each point in the timeseries.

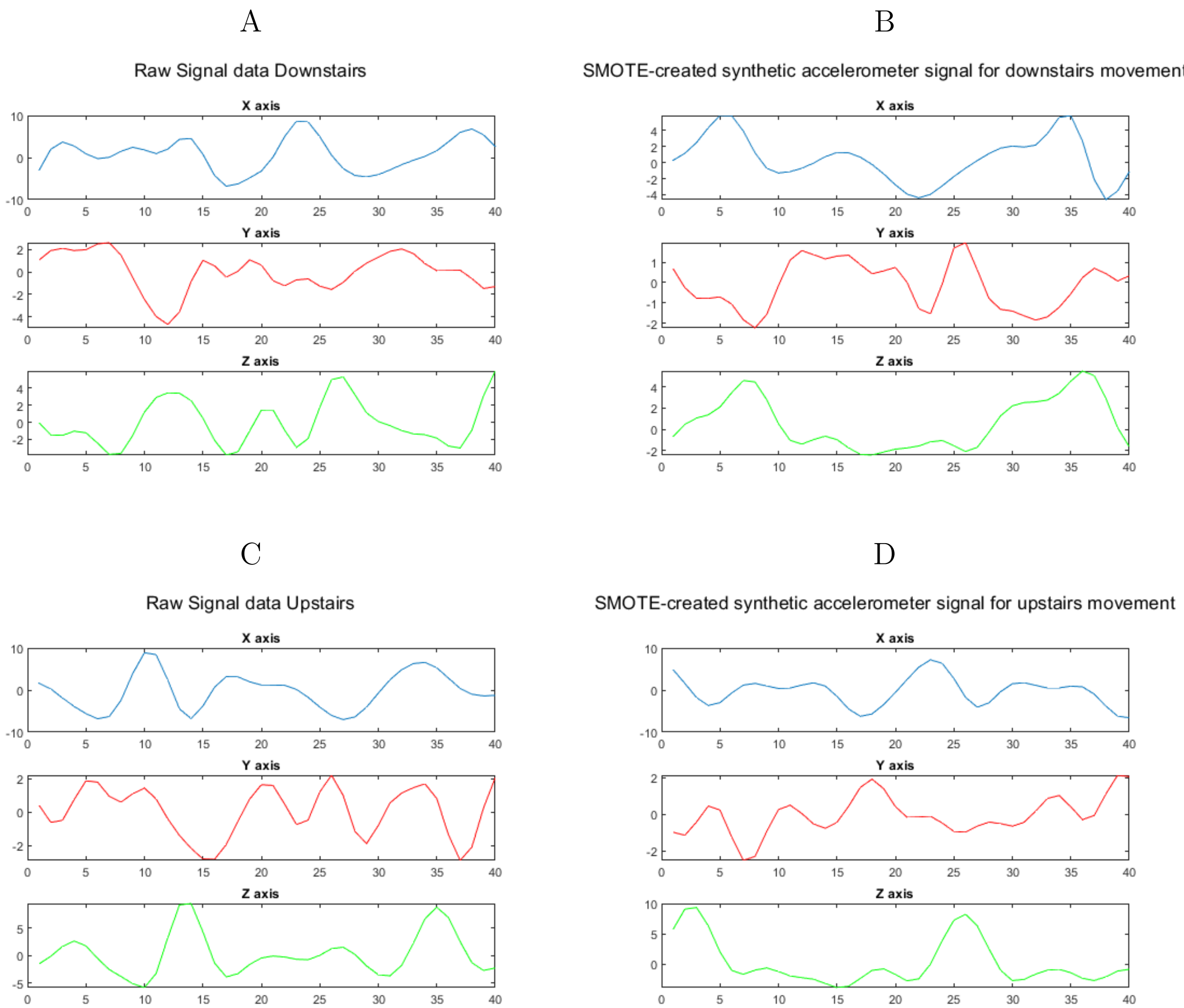


Figure 6.3 Timeseries comparison of timeseries and SMOTE triaxial accelerometer signals. A and C represent random samples of unmodified downstairs and upstairs activity respectively, while B and D represent randomly selected SMOTE-generated samples of downstairs and upstairs activity respectively.

6.2.3 Validation and Terrain Resolution

A number of different validation techniques are carried out in this chapter. In general, most experiments are validated using stratified 5-fold cross validation using all participants activity combined, meaning that the data is randomly partitioned into training and testing datasets, repeated 5 times with different combinations of training and testing data. The stratification ensures that there are proportional balances of classes in both the training and testing partitions. Though 10 is generally the standard value of the K in cross validation in standard HAR papers, it was found that training times exceeded 10 minutes of training time per fold, and therefore would have consumed double the amount of time should 10-fold have been used. 5-Fold is generally considered to be the minimal acceptable level of K for machine learning validation (Marcot and Hanea, 2021). By extension, Leave-One-Subject-Out (LOSO) validation was impractical for most of the experiments (Fushiki, 2011). LOSO validation involves training a classifier on all but one of the participants, test the classifier on the data from the remaining participant, and repeat the process for the next participant until all participants have been separately tested. This validation is used in Experiment #4 where the training and testing sets are reduced.

In hyperparameter tuning (Experiment # 2), there is an additional 5-fold nested cross validation (Fearn, 2010). An illustration of the process is depicted in Fig. 6.4. For outer k_o folds, the data is split into training and testing. The training data is further split into an inner sub-fold k_i to tune hyperparameters via k-fold cross validation, with both the inner and outer loop values for k set to five. Thus, for each k_o fold, a set of optimal parameters are obtained, and the classifier is trained with these parameters. The classifier is validated with the test-set data to get an accuracy, and the overall reported accuracy is the aggregated mean value of test accuracy across the k_o folds. While LSTM used 5-folds in the outer cross validation, due to long training times, the internal validation process for parameter tuning in the LSTM instead used a holdout of 25% of the training data. Nested cross-validation is

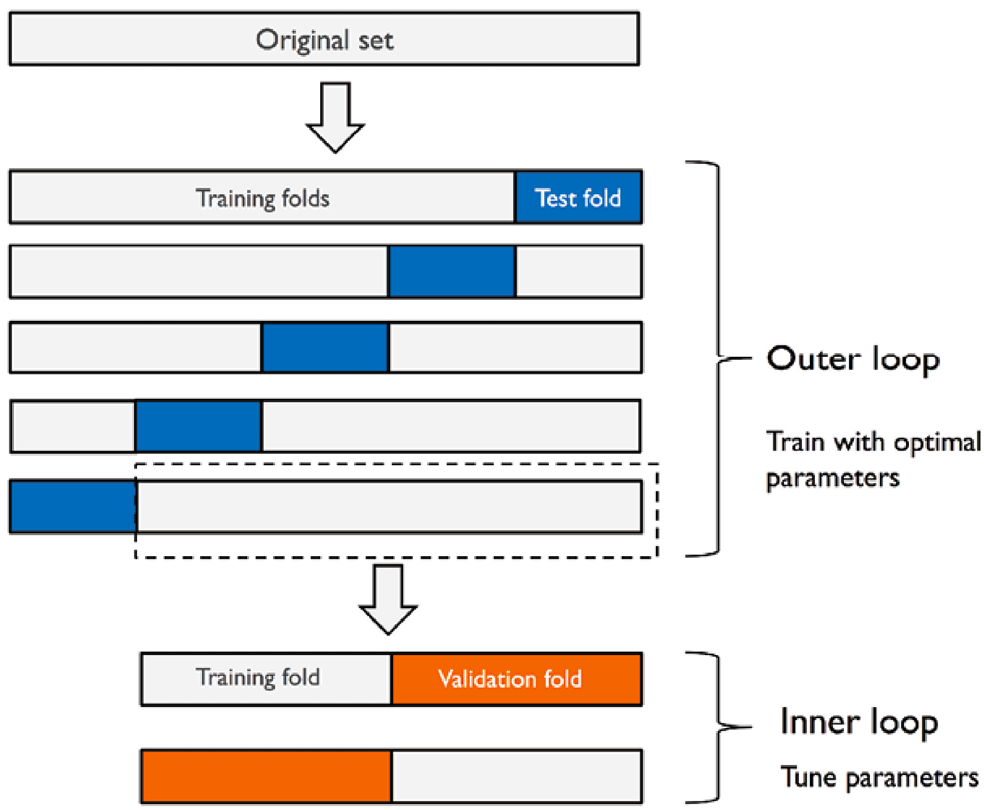


Figure 6.4 Diagram of nested cross-validation, created by (Raschka, n.d.). The image has a CC-BY 4.0 license allowing free distribution of the image.

not used in other experiments, as it significantly increased the computational costs and from some initial experimentation was found that have minimal impact on performances.

In general, most validation metrics used are the mean testing accuracies in percentages, using the proportion of correctly guessed observations to the total amount of observations, to give summative evaluation of the various classification techniques and allows for intuitive cross-comparisons. However, classification accuracies do not always give a truly representative value of the classifiers performance, especially if there is an imbalance in the testing performance; for example, if a binary classifier was trained to only predict class X, and was tested on data that contained 99 X classes and 1 Y class, the classifier would be 99% accurate. In Experiment #2, F1 scores of the individual scores are also calculated for discussion. The F1 score is an alternative metric that evaluates performance based on its balance between the precision and recall, and is defined as follows:

$$F_1 = \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{6.1}$$

Experiments in this chapter also adhere to levels of terrain resolution, as previously indicated in Chapter 5.5.2. The stages are described as follows:

- **Level 1: No Terrain Resolution (5 classes) -** All labels are converged into five activities: Flat walking, Uphill movement, Downhill movement, Upstairs movement, and Downstairs movement.
- **Level 2: Hard/Soft Terrain Resolution (8 classes) -** Terrains are resolved into whether the terrain involved is hard or soft. Hard terrains included concrete, stone and gravel, while Soft terrains included grass, sand and red ash (a type of clay pitch).
- **Level 3: Full Resolution (24 classes) -** All labels use their full terrain resolution.

Experiments #1, 2 and 4 are all carried out at level 1 of terrain resolution (no terrain used in labels), which has the fastest training times and the terrain resolution is not the focus of those experiments. Experiment #3 uses all 3 levels.

6.2.4 Experiment 1: Feature Selection Process

The first experiment attempted to find which method of feature selection or dimensionality reduction was the most appropriate method of condensing the number of hand-crafted features. This experiment is conducted at Level 1 of Terrain resolution, with a total of five activities (see 6.2.3). With 243 dimensions, it was important that the dimensionality of the dataset would need to be reduced to cut down on processing time in subsequent experiments. The most practical and method of validating the feature selection method is through analysing classification performance, which in turn requires passing the modified dataset to a classifier. This created an additional layer of complexity to the optimization process, as the classifiers themselves require their own tuning. A potential solution to this approach was via a wrapper-based feature selection approach, however the wrapper approach is exhaustive and requires considerable numbers of training and validation iterations to reach global minimization of the classifier's training loss function. This drawback became apparent when employing the SFS wrapper feature selection on the dataset; even with undersampling and holdout cross-validation to try and reduce computation costs, the size and dimensionality of the dataset was too large such that SFS was unable to reach a solution for a single classifier after 8 hours of calculation with 16GB of RAM and 3.10GHz processing frequency - irregardless of whether Forward or Backward SFS were used. Thus, it was considered impractical to try and use a wrapper-based approach.

Instead, the experiment used a heuristic approach where for each tested classifier, the default hyperparameters generated by Matlab in their "Classification" companion app were used as the basis for the feature selection validation process. The default parameters are described in Table 6.1. Through these classifiers, various dimensionality reduction and feature selection processes were employed. To reduce the complexity of the results, only one filter and dimensionality reduction technique were chosen for analysis. The preliminary experiments detailed next document how the feature selection processes were chosen.

As explained in Chapter 4.3.6 there are three inbuilt filter-based feature selection methods for classification: the chi-squared feature selection process, the Relief-F selection process, and the mrMR selection process. To pick the most appropriate filter method, a small experiment was conducted: the full dimensionality dataset was reduced to the top 50 features as ranked by each of the filter feature selection methods. Each of the six supervised classifiers (as chosen in Chapter 4.3.7) were configured with the default parameters from Table 6.1 and then validated using 5-fold cross validation of the entire dataset. The resulting mean accuracies are shown in Table 6.2, demonstrating that mrMR was the most consistent filter selection method in terms of the average classifier performance. The other filter methods were also problematic: Relief-F required very long calculation times (approximately 800 seconds per iteration) and the chi-square score of feature importance generated infinite values, thus features that have infinite feature scores could not be ranked against each other. For the experiment, it was decided to investigate how the classifiers performed with different numbers of top “n” features as ranked by mrMR: 10, 20, 50, 75 and 100.

Table 6.1 Classifier training parameters used in experiment #1 (Level 1 Terrain Resolution)

Classifier	Classification App Preset	Parameters Used
SVM	Linear SVM	Kernel: Linear | Kernel Size: automatic | Box Size: 1 | Coding: One vs. One
kNN	Medium KNN	Distance Metric: Euclidean | Number of Neighbours: 10 | Distance Weight: Equal
RF	Ensemble (Bagged Trees)	Maximum Number of Splits: (total number of observations – 1) | Number of Learners: 30
AB	Ensemble (Boosted Trees)	Maximum Number of Splits: 20 | Number of Learners: 30 | Learning Rate: 0.1
NB	Gaussian Naïve Bayes	Distribution for numeric predictors: Gaussian
Discriminant Analysis	Linear Discriminant	None

SVM is trained as a series of binary classifiers, while all other classifiers are trained as multi-class classifiers, see 6.2.5 for further explanation

Table 6.2 Comparison of different filter feature selection methods for top 50 chosen features (Level 1 Terrain Resolution)

Classifiers	SVM	kNN	RF	AB	NB	LDA	**Mean**
Chi-square	69.16	69.75	68.63	60.96	48.95	67.82	**64.21**
mrMR	70.37	72.80	70.00	60.10	62.35	68.68	**67.38**
Relief-F	67.95	73.48	69.06	55.93	60.20	66.32	**65.49**

In addition to the previously listed feature selection processes, PCA was also considered. To recap the PCA operation (described in detail in Chapter 4.2.5.2.1), the PCA process transforms the dataset into a series of eigenvectors in descending order of the proportion of systematic variance. The resulting output is a square coefficient matrix of size $(n_f * n_f)$, n_f being the number of features. Like mRMR, there are no rules on how many principal components are conserved; if the total variance of the chosen principal components is less than 50% of the total systematic variance, then the compressed dataset will have lost some information on the variance between the classes. Generally, the upper accepted limit to the percentage of variance allowed is 95%, beyond which any further principal components are likely to be modelled on noisy factors (Bro and Smilde, 2014). When PCA was applied to the entire dataset, the first component comprised only 28% of the total systematic variance, and the number of components needed to describe 95% of the total variance was 68 as seen in Fig.6.5. Thus, 2 strategies for PCA were investigated; the first was to keep 68 components to explain 95% of the variability, while the second was to keep all components that explained more than 1% of the variance (around 20 components, conserving 42% of the total variance). Validation was obtained as the mean classification accuracy from 5-fold cross validation applied to the entire dataset. The result is shown in Table 6.3. Evidently, conserving components equivalent to 95% of the variance resulted in significantly better classifier performance, with the exception of the RF classifier. This first approach was used in all subsequent experiments.

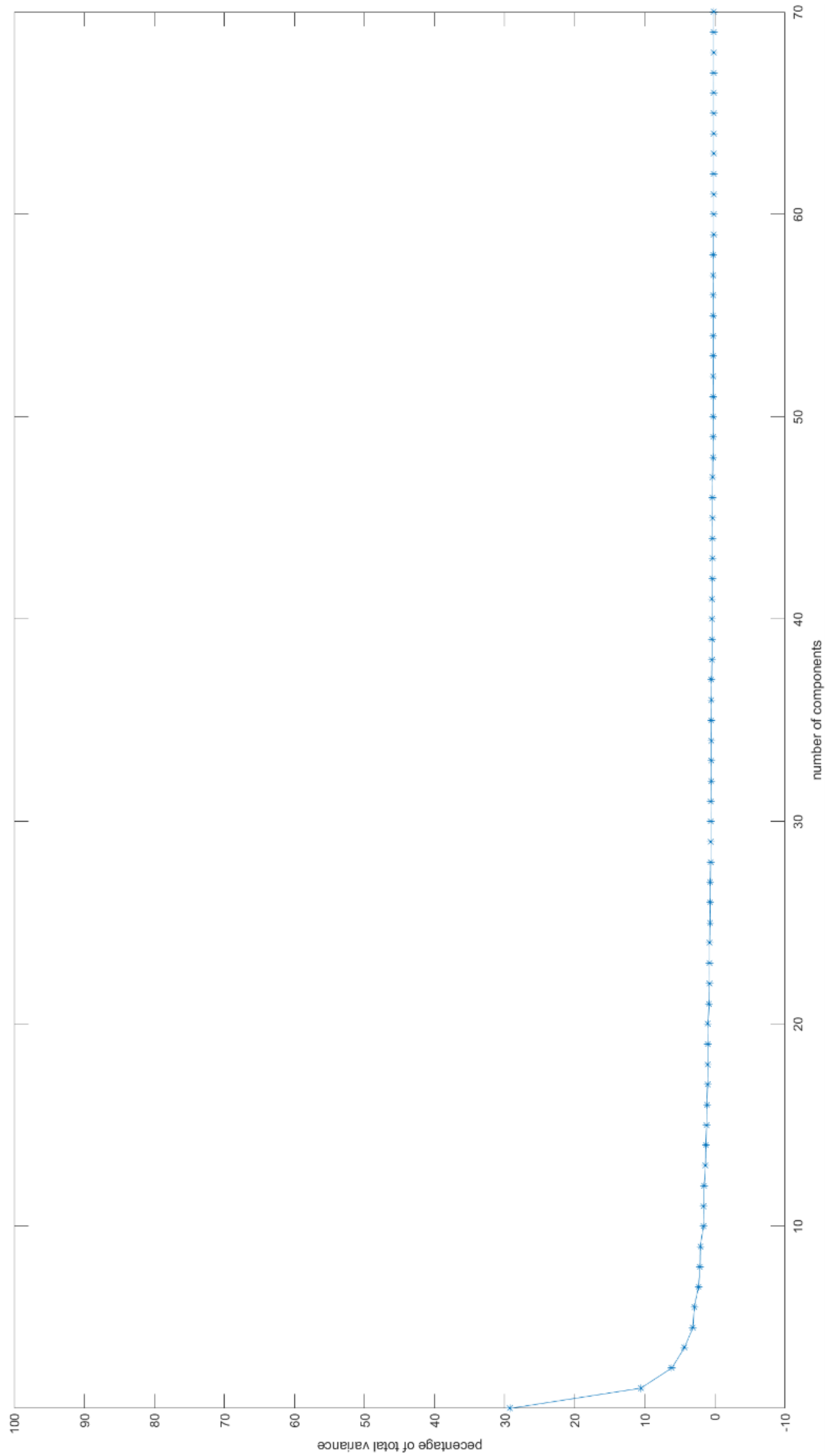


Figure 6.5 Plot of number of PCA components versus the percentage of the total systematic variance each principal component (in descending order) contains

Table 6.3 Comparison of PCA component preservation methods

Classifier	SVM	kNN	RF	AB	NB	LDA	**Mean**
PCA Method #1 (conserve sum of components >95% total systematic variance)	73.70	78.06	68.62	59.11	67.06	71.53	**69.68**
PCA Method #2 (conserve all components >1% total systematic variance)	66.50	72.89	69.13	58.45	63.04	65.13	**65.85**

While Chapter 4.2.5.2.2 explained other non-linear dimensionality reduction process such as kPCA, tSNE and UMAP, these were not considered for this investigation. This is because they are parametric processes and so would require additional optimization of their parameters, whereas with PCA the only parameter to consider are the number of components to conserve. The non-linear dimensionality reduction processes are analysed in Chapter 7.

In addition, this experiment looked at how the combination of mRMR and PCA had an impact on classifier performance. This was varied for different numbers of mRMR-selected features ($10 - 100$) using PCA method #1 (conserve the sum of components that total 95% of the total systematic variance). Applying no dimensionality reduction or feature selection was used as a baseline measurement. Using the LSTM and its inherent feature learning process, the experiment also investigated how using only the raw data from the accelerometer had an impact on performance. Two separate models were trialled: the first used One-Dimensional data acquired from the magnitude of the triaxial accelerometer readings, while the second used each axis of the accelerometer as a separate channel input. For the sake of parity, the standard feature selection and/or dimensionality reduction processes of the 243-dimension feature set are also applied to the LSTM. The two LSTM hyperparameters analysed in this investigation: dropout factor and number of hidden units, were given values of 0.5 and 225 respectively for fast training times. All classifiers were validated using 5-Fold cross validation of the entire dataset.

6.2.5 Experiment 2: Classifier Optimization

The objective of experiment #2 was to identify the best classifier for supervised classification of walking activities. Once the ideal feature selection process was found in the previous experiment, the six tested supervised classifiers and LSTM network underwent hyperparameter tuning to determine the best classifier. The list of eligible hyperparameters for tuning are listed in Table 6.4. In SVM optimization, the Kernel Scale and One vs. All parameters in SVM models were excluded from analysis as inclusion of those parameters as variables resulted in training times exceeding an hour per optimization iteration. The SVM thus trains through the One vs. One strategy, performing binary classification in pairs of classes then selecting the class with the most amount of "votes" from all decision rounds. All other classifiers are trained as multi-class classifiers.

Table 6.4 Parameters tuned in experiment #2

Classifier	Parameters Tuned
SVM	Kernel Type, Box Size
KNN	Number of neighbours, Distance Metric, Distance Weight
RF	Maximum number of learners, maximum number of splits
AB	Maximum number of learners, maximum number of splits, learning rate
NB	Type of distribution, Type of kernel (for kernel distribution)
Discriminant Analysis	Type of discriminant analysis (Linear, Quadratic)
LSTM	Dropout factor, Number of Hidden Units

For the supervised classifiers, the nested tuning process (see Section 6.2.3) was handled via Matlab's Classification companion app. The procedure utilised a Random Search algorithm with 20 maximum iterations (this excludes NB and LDA classifiers which had less than

20 possible input combinations)(Bergstra and Bengio, 2012). A Random Search algorithm will, for each iteration, select hyperparameter values at random until the training error decreases, making the training process less exhaustive and time consuming compared to grid search. Bayesian optimization was the other available option for tuning, and while it can reach optimization in fewer numbers of training iterations compared to grid and random search (Shahriari et al., 2016; Wu et al., 2019), the process is not recommended when the dimensionality of the dataset is greater than 50, thus was not viable for the high dimensionality dataset (Mathworks, n.d.[a]). After the best parameters have been determined, they are relayed to the input of the classifier a final time and validated on the untouched test partition of data, the mean accuracy is saved and the entire process is repeated a further 4 times.

The LSTM network, due to its long training times, used a simple grid-search approach with discrete values for the dropout factor (0.1 to 0.5) and the number of hidden units (175 to 350). The hidden unit scale was determined using a crude formula recommended on the Stack Exchange forums (Hobs, n.d.):

$$N_h = \frac{N_s}{(\alpha * (N_i + N_o))} \tag{6.2}$$

The number of hidden neurons, N_h, can be roughly determined through dividing the number of input samples N_s by the sum of the number of input and output units N_i and N_o, and multiplied by a scalar factor α, which is recommended to range between 5 and 10. Thus, the scale was determined by changing the value of α in units of 1 and recording the corresponding number of hidden units. As stated in Section 6.2.3, the LSTM uses a holdout of data for grid-search optimization, the best parameters are then configured to the LSTM and validated with another holdout not used for optimization to obtain the mean accuracy, which is stored and the process repeats a further 4 times.

For a sanity check, this experiment is repeated with all classifiers using a single fold of validation (essentially holdout validation) to ensure that the resulting validation metrics obtained were consistent, but is not aggregated into the main results. The experiment is conducted at Level 1 of terrain resolution, and is validated with mean recognition accuracies to reflect the overall performance of each classifier.

6.2.6 Experiment 3: Label Terrain Resolution

In this experiment, the appropriate level of terrain resolution for supervised HAR is determined. As previously speculated in Chapter 5.5.2, there was an expected trade-off between the resolution of detail in the activity and the corresponding classifier performance. This experiment aimed to show at what specific label of terrain resolution can an acceptable level of classifier accuracy be obtained. Using the best performing classifiers from experiment 2 alongside their best performing hyperparameters, the terrain resolution was increased in stages (see Section 6.2.3 for reference). Each level of resolution is once again tested using 5-fold cross validation of the entire dataset.

6.2.7 Experiment 4: Subject Cross-Validation

In the three preceding experiments, all validation used 5-fold cross validation of the entire dataset, meaning all twelve subjects' feature data is combined into a single matrix and then split for training and testing. In the final experiment of this chapter, the objective was to determine whether a classifier trained exclusively on non-amputated or amputee subjects would be robust enough to detect physical activities in a single amputee subject. Hence, this experiment uses Leave One Subject Out (LOSO) validation. There were two methods of training the classifier:

- **Method 1: Non-Amputee Subject Training** For the best performing classifiers, train data on the 8 non-amputated subjects, then use LOSO validation to test the classification accuracy for each of the 4 ILLA subjects individually.

- **Method 2: ILLA Subject Training** All data from non-amputated subjects are ignored. Leave-One-Subject-Out validation is again performed with the best classifiers, this time training the data on 3 amputee subjects and testing on the remaining amputee subject, repeating the process for each amputee.

In this experiment, the terrain resolution is kept at level 1.

6.3 Results

6.3.1 Experiment 1: Feature Selection Process

Table 6.5 demonstrates the accuracies of the classifiers through various feature selection and dimensionality reduction methods. Four out of the six supervised classifiers performed best when no feature selection was applied, however the LDA classifier failed to function, with accuracy equivalent to a random guess. Two of the other classifiers performed best with just PCA applied. The LSTM demonstrated good classification performance regardless of whether the input to the classifier was the feature set or raw data under various dimensionality reduction processes. With 3 dimensions of raw triaxial data, the LSTM had comparable performance to the best overall performing classifier (kNN with no feature selection), and so all subsequent experiments with the LSTM utilised 3D timeseries data. Although mrMR-based feature selection with 75 features had the highest overall performance, it was decided to proceed with using PCA as the dimensionality reduction method for the other six supervised classifiers. The LDA and NB classifiers both performed the best under PCA dimensionality reduction, while no classifier performed the best with the mrMR filter feature selection. The combination of filter and PCA methods generally appeared to underperform, with only the selection of 100 features followed by PCA being the only method to exceed 70% classification accuracy. The most reasonable explanation for this is that as the supervised feature selection directly chooses features that benefit classification, the PCA effectively filters out

Table 6.5 Mean accuracy of supervised classifiers (%) under various dimensionality reduction methods. (Level 1 Terrain Resolution)

Classifier	SVM	kNN	RF	AB	NB	LDA	LSTM	Mean
No Dimensionality Reduction	**76.28**	**78.43**	**74.03**	**64.16**	57.81	22.35	72.76	63.69
Filter (10)	57.74	53.62	56.12	54.34	56.03	56.57	51.80	55.17
Filter (20)	64.94	65.66	63.73	56.75	61.07	63.17	60.86	62.31
Filter (50)	70.67	74.41	70.90	61.49	63.62	67.52	67.43	68.01
Filter (75)	73.52	76.91	72.59	62.68	62.86	71.15	75.73	**70.78**
Filter (100)	73.57	76.23	72.41	62.71	61.87	19.05	76.08	63.13
PCA	73.70	78.06	68.62	59.11	**67.06**	**71.53**	76.73	70.69
Filter (10) + PCA	54.73	52.25	52.36	48.67	53.41	54.15	50.25	52.26
Filter (20) + PCA	64.67	64.14	63.00	57.74	61.87	63.30	62.87	62.51
Filter (50) + PCA	69.43	72.91	66.82	62.05	64.98	67.74	70.58	67.79
Filter (75) + PCA	71.33	75.82	67.98	61.81	65.78	69.10	73.39	69.32
Filter (100) + PCA	71.92	76.64	68.75	61.77	65.23	70.50	75.61	70.06
Accelerometer magnitude timeseries data (1D)	N/A	N/A	N/A	N/A	N/A	N/A	72.16	N/A
Accelerometer triaxial timeseries data (3D)	N/A	N/A	N/A	N/A	N/A	N/A	**78.00**	N/A

"Filter (n)" indicates top n features as chosen by mrMR are preserved || N/A = not applicable || 1D = one dimensional data || 3D = 3 dimensional data

low variance aspects of the feature set that could potentially benefit the classifier, thereby causing a reduction in classification accuracy. When 100 features are chosen through feature selection, most of these features are likely to be noisy features with minimal variance, which become compressed with PCA giving it comparable performance to standalone application of PCA.

To highlight the main findings: the six non-LSTM classifiers would use PCA as the dimensionality reduction process in Experiment #2 onwards, while the LSTM would use the triaxial timeseries accelerometer data, using feature learning to automatically derive the features in the timeseries signal for classification.

6.3.2 Experiment 2: Classifier Optimization

Excluding the LSTM model, which used 3D raw data, each of the six classifiers had PCA applied on the training dataset, the coefficients of which were applied to the testing dataset to transform into eigenvalues. The results are shown in Table 6.6, and from these the "best" classifiers were identified as the SVM and LSTM models. Neither classifier statistically significantly improved mean classification accuracy over the third-best performing classifier (kNN); t-test values demonstrated the SVM and LSTM had statistically insignificant difference in means with $p = 0.2$ comparing the SVM to kNN and $p = 0.14$ for the LSTM to the kNN. However the kNN and the remaining 4 classifiers were not used in subsequent experiments as some preliminary implementations of those classifiers were found to have negligible or weaker classification performance than the SVM. The tuning time for the inner cross-validation processes are also included in the table, with some contrasting tuning times; for example the RF tuning process is 45 times faster than the SVM process. The differences are explained by how the hyperparameters affect the classification performance; in the SVM, difference of the kernel and box constraint likely has negligible difference on performance, therefore it takes the random search algorithm longer to find the right set of parameter combinations to boost performance. The LSTM had the longest tuning time, requiring 50 minutes of optimization for a single iteration; this further highlighted the inability to perform 10-fold validation, as this would require 8.45 hours of tuning with the potential risk of the application crashing and losing all results.

While the LSTM classifier had slightly higher recognition accuracies, the SVM was considered a comparable classifier due to its lower standard deviations in the testing folds. The best parameters indicate the hyperparameters that were present in the highest scoring fold of the outer cross validation.

Table 6.6 Comparisons of supervised classifiers under 5-fold cross validation, the time required for tuning the classifier (per iteration) and the best suited hyperparameters for each (Level 1 Terrain Resolution)

Classifier	5-Fold Accuracy (%)	Fold Std. (±%)	Mean hyperparameter tuning time (s)	Best Parameters
SVM	77.22	0.54	1667.6	Kernel: Gaussian | Box Constraint: 193
kNN	75.76	2.26	632.3	Num. Neighbours: 12 | Distance Metric: Correlation | Distance Weight: Squared Inverse
RF	71.84	0.69	37.0	Num. Learners: 188 | Num. Splits: 4392
AB	69.23	3.18	948.5	Num. Learners: 211 | Num. Splits: 910 | Learn Rate: 0.07
NB	64.40	0.95	1145.1	Gaussian Kernel Distribution
LDA	73.91	0.93	18.15	Quadratic
LSTM	78.46	2.89	3042.7†	Dropout Factor: 0.2 | Num. Hidden Units: 190

† Different tuning methodology: tuning time represents total time to optimize LSTM classifier through grid search for a single hold-out partition of the data

Overall, hyperparameter tuning appeared to have minimal impact on improving classifier performance, compared to accuracies reported in Table 6.5 where default parameters were used. In fact, the accuracy of the kNN and NB models decreased compared to their results in experiment 2 when PCA was performed. There are two factors which explain this: as the cross-validation process did not use the same indexes for partitioning training and testing data between the experiments, the partitions in the first experiment may have given the classifiers a more favourable performance due to chance. The second explanation is, since the hyperparameter tuning must be performed in a sub-fold of the partition, the classifier is working with a smaller subset of the data than they are in the first experiment. Thus, the parameters which best satisfy the inner validation process may not generalize to the outer validation process. This also likely means that the choice of hyperparameters in these models are inconsequential in terms of their ability to improve classification performance. Naturally, the hyperparameters could still make an adverse impact on classification performance if chosen poorly.

6.3.3 Label Terrain Resolution

All subsequent experiments used SVM and LSTM models with their best parameters from Table 6.6, and no further hyperparameter tuning was performed. An exception to this was the number of hidden units used in the LSTM model: as the training set size increased or decreased due to performing SMOTE on the new classes, the number of hidden units was dynamically changed using the approximation formula from Eq. 6.2.5 with the α scaling factor set to 9, which corresponded the optimal number of hidden units in the LSTM. While the accuracies reported in experiments 3 & 4 may be theoretically lower that what was possible with hyperparameter tuning, the results in experiment 2 demonstrated that these would be unlikely to have a significant impact.

The mean classification accuracy results from varying terrain resolution can be seen in Table 6.7, which is accompanied with F1 scores for each activity in Tables 6.8-6.10. In the base terrain resolution, flat and stair data had notably greater F1 scores compared to hill data, which is to be expected due to stair movement having significantly different kinematic behaviour from standard walking (Park et al., 2019). Uphill movement typically had less confusion than downhill movement, which is supported by Kimel-Naor, Gottlieb, and Plotnik (2017) whose findings suggest that downhill movement and flat movement are often very similar due to minimal required changes in gait kinematics for downhill movement.

Table 6.7 Comparison of mean classification accuracy in SVM and LSTM classifiers for varying resolutions of activity label, using 5-fold cross-validation

Label Resolution Level	1	2	3
No. of Classes	5	7	24
SVM accuracy(%)	77.22±0.54	62.60±2.73	32.74±2.46
LSTM accuracy(%)	78.46±2.89	73.77±1.83	41.85±4.19

The mean accuracy of both classifiers decreased as terrain resolution (therefore the number of activities) increased. While the drop-off in mean classification accuracy for the LSTM model was less than in the SVM model, the F1 scores at each stage of terrain resolution (Tables 6.9-6.10) reveal that that many of the individual classes for Levels 2 and 3 have poor F1 scores, and the SVM model had larger F1 scores in many class-by-class comparisons. By viewing the confusion matrices of the LSTM model (Figs. 6.6-6.11), the higher accuracies are a result of having high recall for the majority class ("Flat", "Hard, Flat", "Concrete, Flat" or "Concrete"), indicating that despite balancing the classes with SMOTE, the LSTM model has high model complexity and has overfit, as it has statistical tendency to simply predict the majority class (Allamy, 2014). The SVM models, while having moderately poor performance, had less underfitting and better precision in the minority classes.

Table 6.8 Mean F1 Scores for Level 1 Label Resolution

Level 1 (No Terrain)			
Activity	**# of Test Samples**	**LSTM F1 Scores**	**SVM F1 Scores**
Downstairs	131	0.800	0.746
Upstairs	138	0.801	0.757
Flat	3826	0.838	0.842
Uphill	670	0.687	0.699
Downhill	701	0.628	0.617

Table 6.9 Mean F1 Scores for Level 2 Label Resolution

Level 2 (Hard/Soft Terrains)			
Activity	**# of Test Samples**	**LSTM F1 Scores**	**SVM F1 Scores**
Downstairs	131	0.803	0.758
Upstairs	138	0.774	0.751
Hard, Flat	3506	0.802	0.823
Hard, Uphill	647	0.699	0.708
Hard, Downhill	668	0.631	0.615
Soft, Flat	321	0.571	0.613
Soft, Uphill	22	0.271	0.362
Soft, Downhill	33	0.276	0.367

Table 6.10 Mean F1 Scores for Level 3 Label Resolution

Level 3 (All terrains)			
Activity	**# of Test Samples**	**LSTM F1 Scores**	**SVM F1 Scores**
Concrete, Camber, Downhill, Parallel	114	0.348	0.380
Concrete, Camber, Downhill, Perpendicular	19	0.037	0.0
Concrete, Camber, Parallel	784	0.508	0.623
Concrete, Camber, Perpendicular	176	0.089	0.044
Concrete, Camber, Uphill, Parallel	108	0.393	0.505
Concrete, Camber, Uphill, Perpendicular	18	0.032	0.0
Concrete, Downhill	510	0.481	0.583
Concrete, Flat	2440	0.424	0.725
Concrete, Uphill	485	0.555	0.695
Downstairs	132	0.715	0.749
Grass, Downhill	30	0.315	0.371
Grass, Flat	269	0.510	0.626
Grass, Uphill	16	0.251	0.291
Gravel, Downhill	9	0.178	0.242
Gravel, Flat	6	0.074	0.0
Gravel, Uphill	1	0.0	0.0
Red Ash, Flat	5	0.174	0.250
Sand, Downhill	2	0.0	0.0
Sand, Flat	47	0.333	0.520
Sand, uphill	6	0.186	0.270
Stone, Downhill	16	0.159	0.094
Stone, Flat	99	0.184	0.194
Stone, Uphill	35	0.383	0.413
Upstairs	139	0.694	0.734

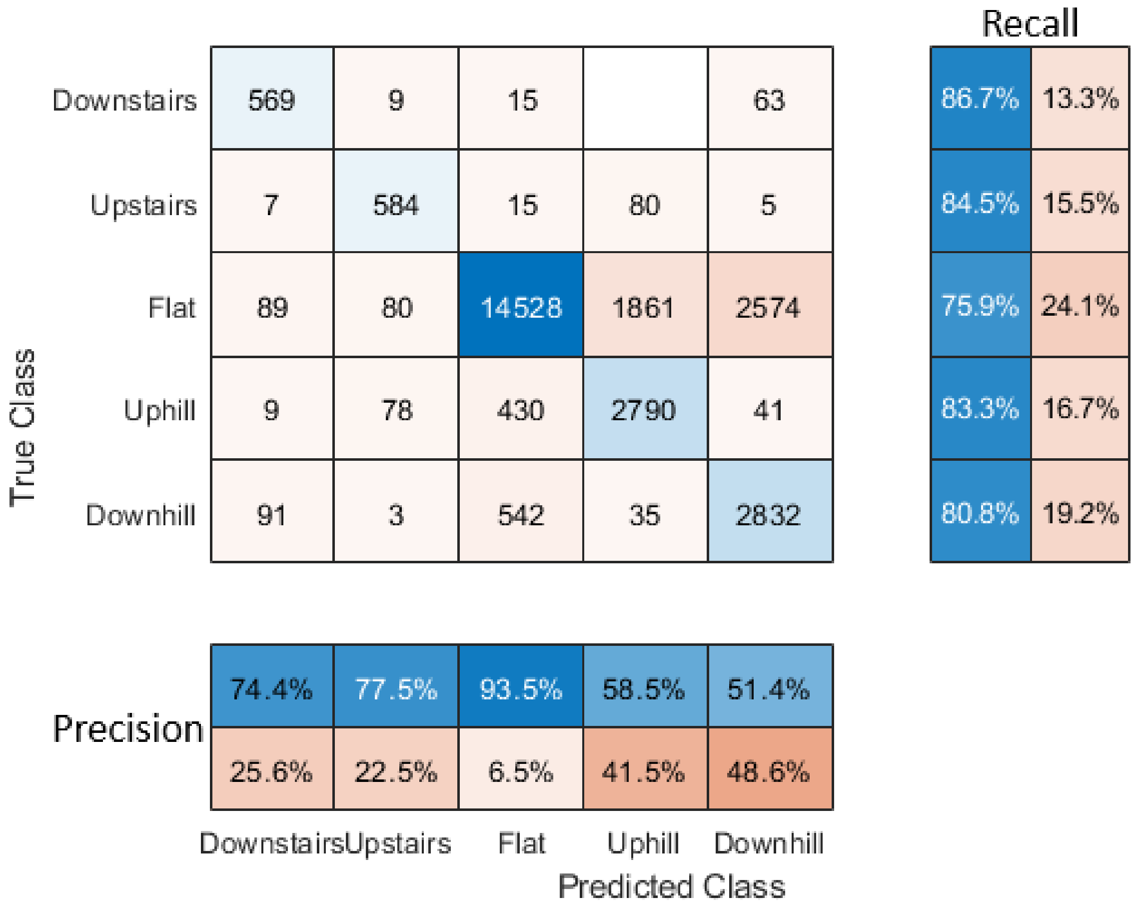


Figure 6.6 Confusion chart of LSTM classifier at level 1 label resolution

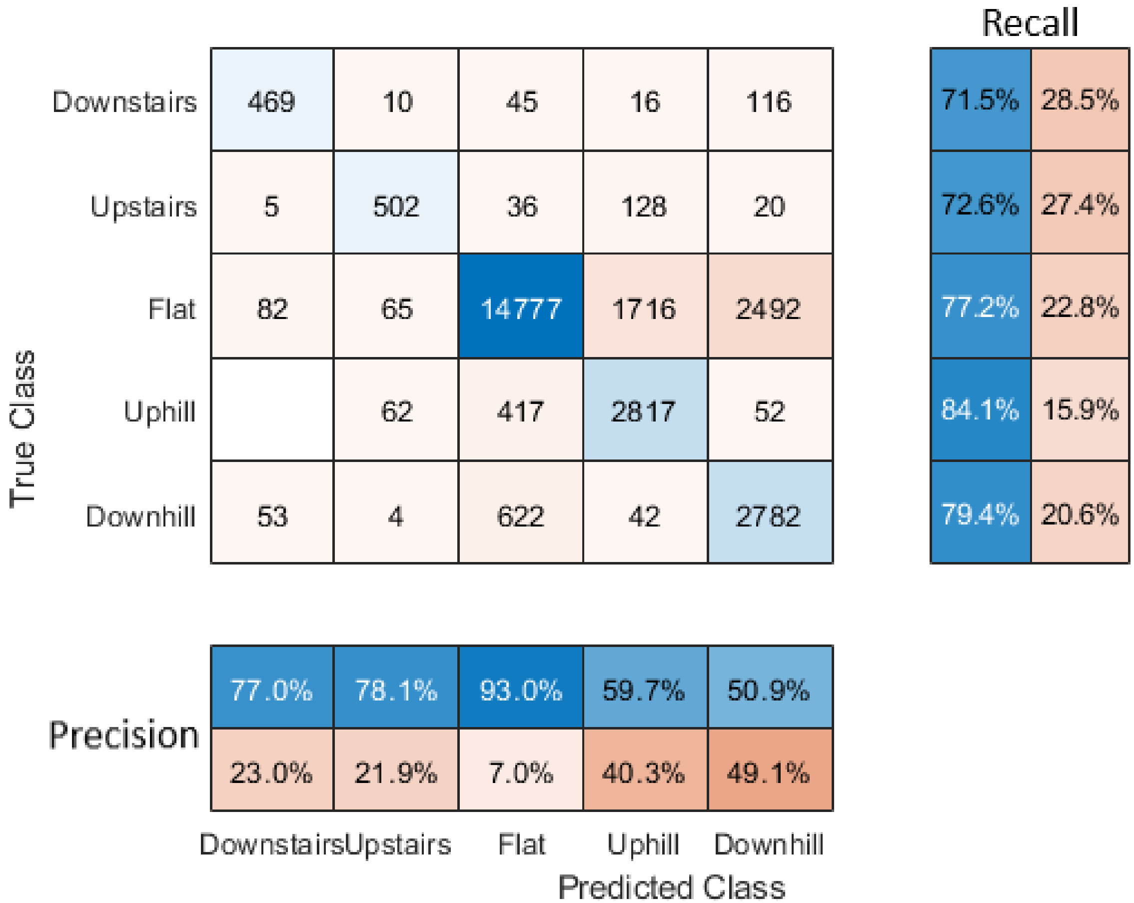


Figure 6.7 Confusion chart of SVM classifier at level 1 label resolution

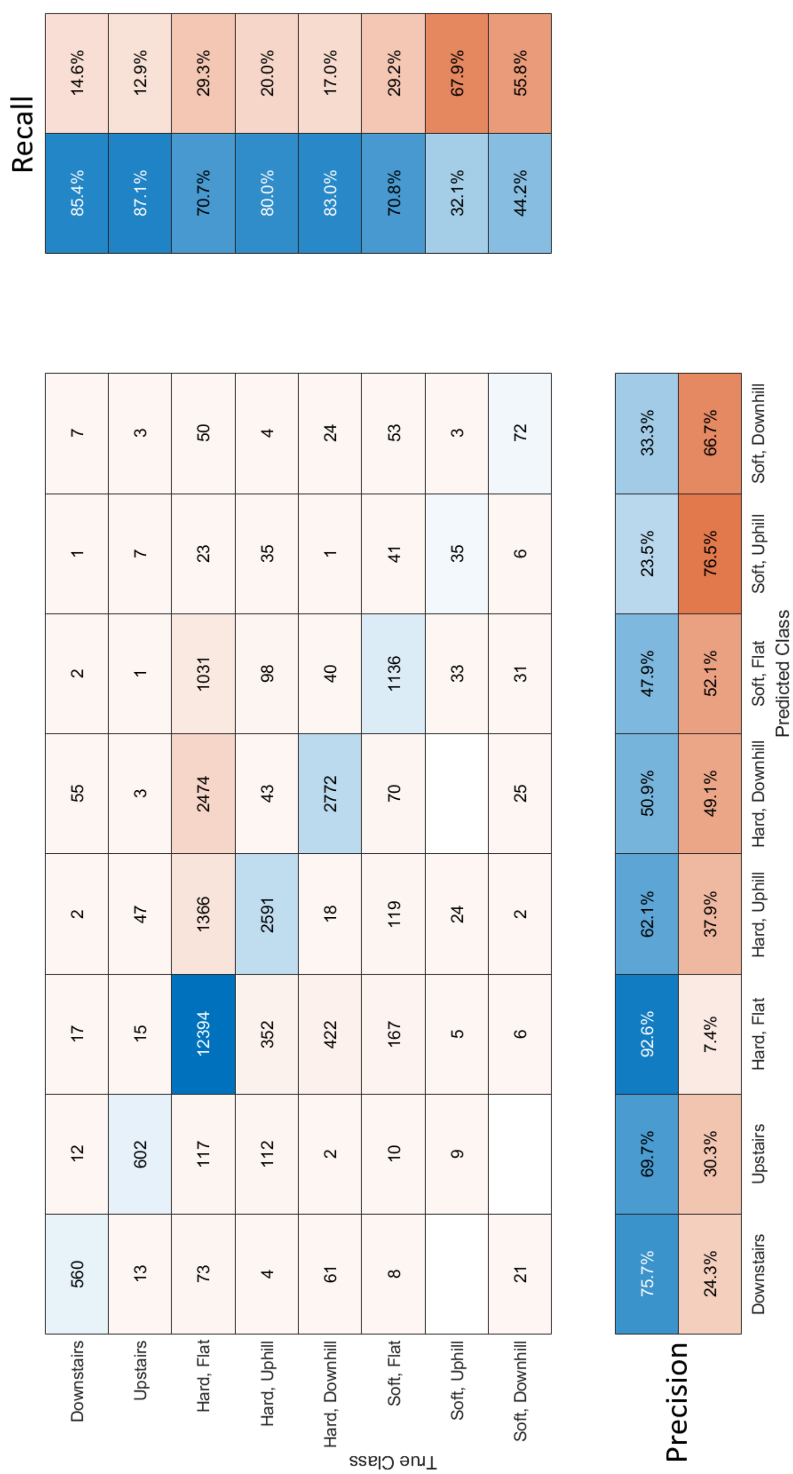


Figure 6.8 Confusion chart of LSTM classifier at level 2 label resolution

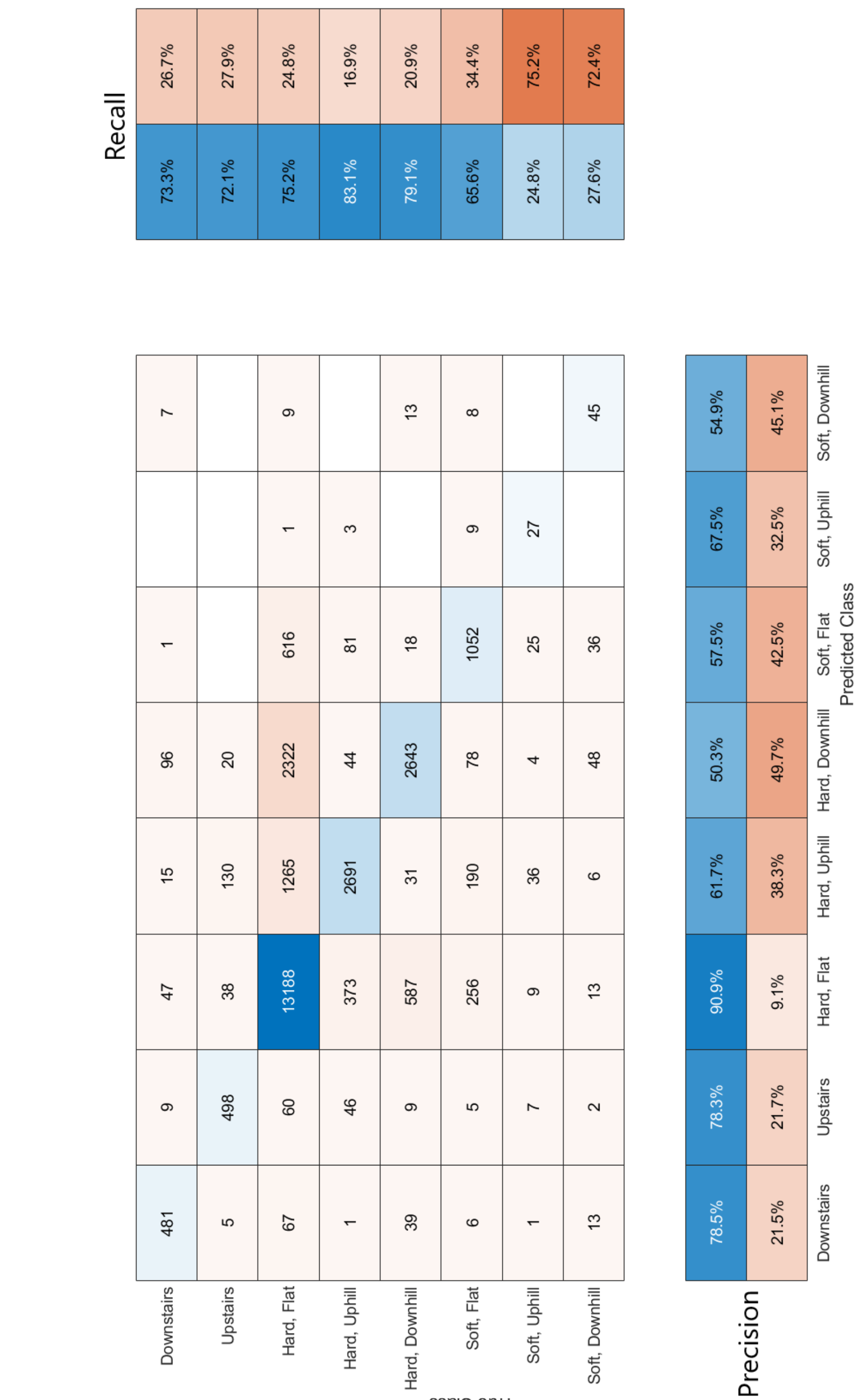


Figure 6.9 Confusion chart of SVM classifier at level 2 label resolution

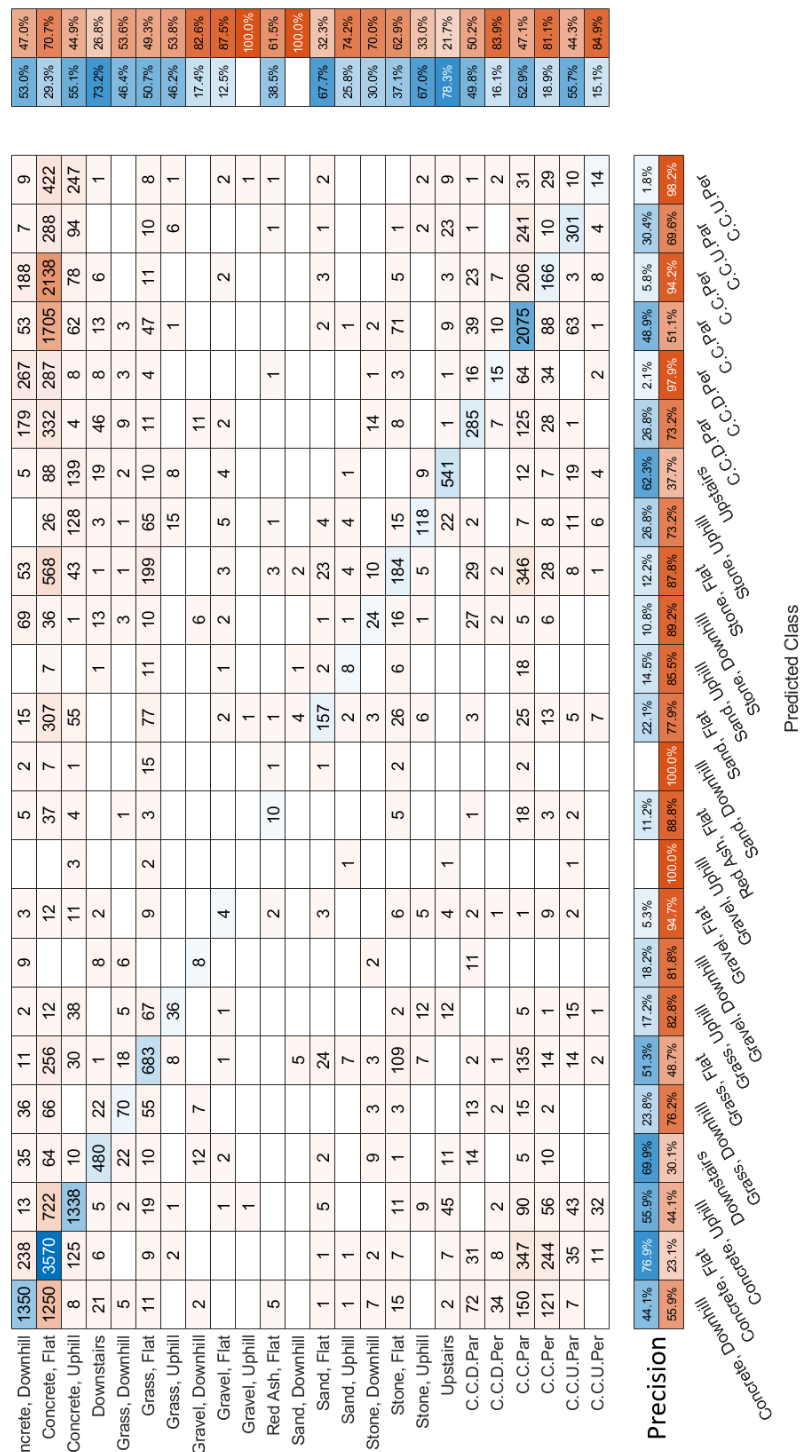


Figure 6.10 *Confusion chart of LSTM classifier at level 3 label resolution

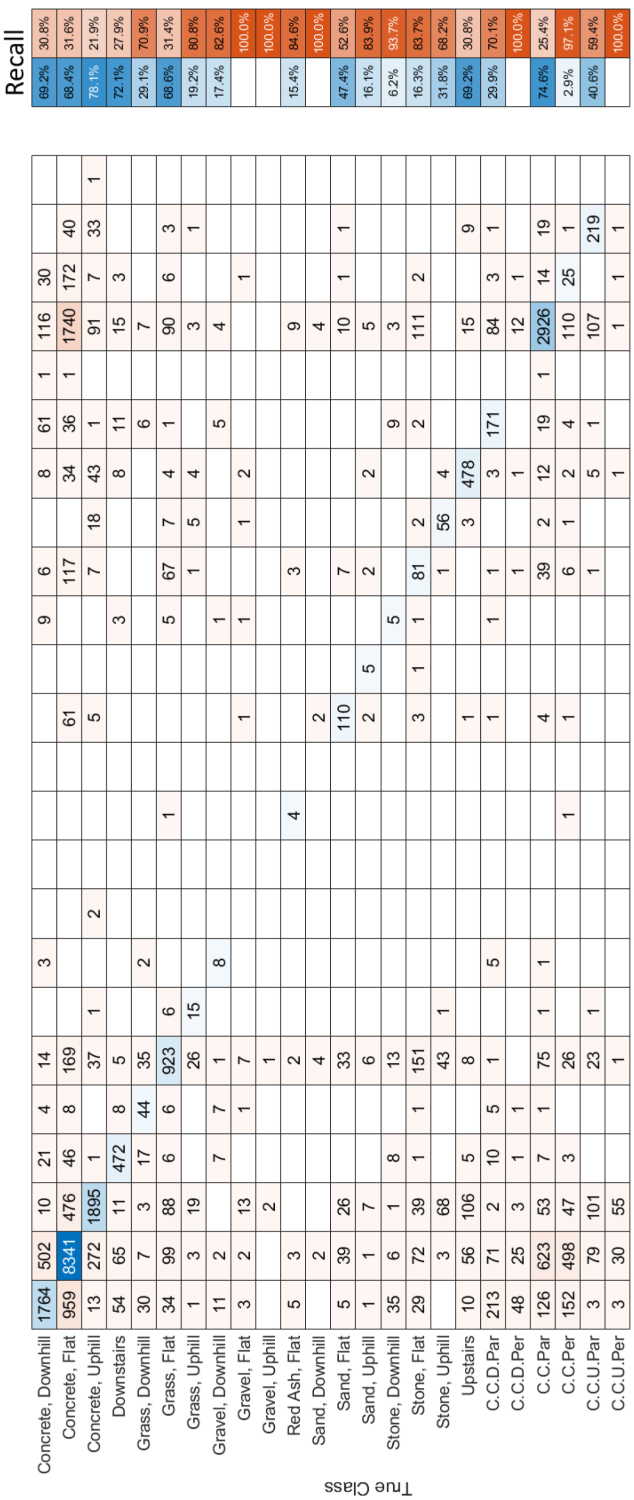


Figure 6.11 *Confusion chart of SVM classifier at level 3 label resolution

* - for Figs. 6.10 & 6.11, certain walking activities are abbreviated as follows:

- C.C.D.Par - "Concrete, Camber, Downhill, Parallel"
- C.C.D.Per - "Concrete, Camber, Downhill, Perpendicular"
- C.C.Par - "Concrete, Camber, Parallel"
- C.C.Per - "Concrete, Camber, Perpendicular"
- C.C.U.Par - "Concrete, Camber, Uphill, Parallel"
- C.C.U.Per - "Concrete, Camber, Uphill, Perpendicular"

6.3.4 Experiment 4: Subject Cross-Validation

Table 6.11 presents the results of training classifiers by both an amputee and non-amputated population and testing on one amputee, using "no terrain" resolution. The experiment was unfortunately unsuccessful; while SVM accuracies are moderately poor, the LSTM models failed completely, for some subjects having a recognition accuracy lower than random guess. Despite the significantly smaller size of training dataset available, the mean classification performance of the SVM when trained by non-amputated subjects was lower than the performance when the SVM was trained by other amputees. A significant confounding factor in this study, as previously described in 5.9.2, was the fact that the amputees walked over considerably more variations in terrain than the non-amputated individuals, and even within the amputee group, some amputees walked on certain terrains that other amputees rarely or never walked over. For instance, amputee subject #1 walked over stony terrain, with no instances of sandy terrain, while the opposite was true for amputee subject #2. The fact that amputee subject #4 was a bilateral amputee while the other three were unilateral may have caused further confusion in the models. The LSTM model had wildly inconsistent inter-subject accuracy and high standard deviations, but given the much smaller size of training dataset, particularly when trained with only amputee subjects, this was not unsurprising.

The confusion charts for each subject (by Classifier and by training method) are located in Appendix C.

Table 6.11 Mean LOSO accuracies of each of the 4 ILLA subjects (level 1 terrain resolution)

Non-amputee Subject Trained Classifiers:						
Subject	A1	A2	A3	A4	Mean	Inter-subject Std.
SVM accuracy(%)	52.41	60.92	46.44	58.41	**54.55**	±5.61
LSTM accuracy(%)	24.97	48.71	14.79	25.60	**28.52**	±12.42
Amputee Subject Trained Classifiers:						
Subject	A1	A2	A3	A4	Mean	Inter-subject Std.
SVM accuracy(%)	58.12	52.99	54.08	61.55	**56.68**	±3.40
LSTM accuracy(%)	53.97	10.26	43.22	16.96	**31.10**	±18.06

6.4 Discussion

6.4.1 Main Findings & Interpretations

The main outcomes of this research are summarized as follows: by analysing walking activity data carried out by ILLAs and non-amputated individuals in free-living settings with supervised classifiers, this research was able to achieve good results, but with significant room for improvement. The best suited classifiers for the supervised approach were discovered to be an SVM which calculated 243 features in the time, frequency, and wavelet domains, with compressed dimensionality via PCA to 68 principal components, and an LSTM Network which required only the raw sequential triaxial accelerometer data. The LSTM model generally had superior mean classification accuracy but required much longer training times and had perceptible underfitting, particularly towards the majority class. Given these significant drawbacks, the SVM was overall the more robust classifier and, if based purely on these results, would be the most suitable classifier in a clinical-based activity monitoring system. The LSTM's design is based on a simple preconfigured model available from Matlab's Deep Learning designer application. In cutting-edge deep learning HAR experimentation, LSTM models usually have more design complexity. Zhao et al. (2017) for instance implemented a bidirectional LSTM, which in comparison to ordinary LSTM learns time dependencies in both directions of the signal, acquiring an average of 94% in F1 scores. Sun et al. (2018) meanwhile used a combination of CNN and LSTM architecture. The CNN component performed feature extraction from convolution of the raw input data, from which the LSTM recognizes the time-dependencies of the CNN-forged features. They perform classification through an Extreme Learning Machine - an alternative to the SoftMax classifier (Huang et al., 2012) - and were able to achieve strong recognition accuracies for high level activities. Given that this machine learning problem was approached from a broad view (meaning to try lots of different techniques and solutions rather than focus on one), the design and subsequent optimization of a more complicated LSTM network was not practical within the scope and timeframe of the thesis. LSTM and other deep learning networks could have been investigated in more detail if it had been decided early on to focus and specialize in the implementation of a deep learning network over shallow classifiers. Technical limitations of the deep learning networks also arose as a result of Matlab's lack of official support for deep learning classifiers; for example, Extreme Learning Machines are not supported in their deep learning classification app and require downloading unofficial community-supported toolboxes in place. Future implementations of the project should utilize Keras in a Python™ (Python Software Foundation, DE, USA) Integrated Development Environment (IDE) for more refined control of deep learning parameters as well as for accessing a wider range of implementable deep learning libraries (Keras, n.d.).

From the terrain-resolution experiments (experiment #3), there is some potential to expand the typical walking activities captured in a HAR investigation to include hard and soft terrains as additional outcomes. While "soft, flat" had acceptable F1 scores, "soft, uphill" and "soft, downhill" did not, this is likely due to having much smaller quantities of these activities compared to "soft, flat" and the hard terrain activities. While the inclusion of SMOTE can help increase performance in these minority classes and prevent overfitting (Qazi and Raza, 2012), if the initial training data size is too small, the SMOTE process will still be unable to successfully capture the full range of potential feature data that minority class can exhibit because it can only create new samples within a nearest neighbour proximity to that existing data, which can be especially problematic if the data is noisy (Flores et al., 2018). Balancing the dataset in the future may require greater degrees of undersampling in the majority class to provide a better initial balance to the data prior to SMOTE. Unfortunately, the terrain resolution beyond "hard/soft terrain" (level 3) appears to be impractical to detect accurately, at least within the scope of the data available for the investigation. There would likely require significantly more data from the minority classes and perhaps require more additional sensors. For example, a second ActivPAL attached to the other leg could use inter-sensor axis correlation coefficients to detect the presence of laterally uneven ground. The presence of "camber movement" in particular may be impossible to detect with consistent accuracies due to its high similarity with "concrete, flat" movement coupled with variations in road designs (for example, having different degrees of slope).

The subject cross-validation experiments (experiment #4) were regrettably completely unsuccessful, even at the most basic level of terrain resolution. This likely was due to the amputees carrying out different walking activities, both from the non-amputated individual group and from each other. Future investigations would need to acquire considerably larger numbers of amputee volunteers with a diverse range of amputation types, and work towards avoiding the gender and age biases shown in this dataset which may have had further influence on the results. As the evidence from the experiment indicates, the clinical implication of experiment #4 is that a classifier for the detection of walking activities in ILLAs cannot be reliably trained on activity data from a population with non-impaired gait, and by comparing the LOSO validation of the bilateral amputee (subject A4) to the others, this experiment also indicates that different classifiers may need to be trained for different types of lower limb amputation.

6.4.2 Comparisons with Other Studies

The recognition accuracies in this investigation were far from ideal: each of the lab-based studies identified in Labarrière et al. (2020) that used only an IMU sensor for data collection had comparable or superior recognition accuracies (Mai et al., 2018; Mai et al., 2018; Stolyarov, Burnett, and Herr, 2018; Su et al., 2019). This study is however does have comparable performances with other free-living HAR investigations. Gyllensten and Bonomi (2011) found that activity data trained in laboratory conditions dropped from an average recognition accuracy of 95.4% to 75.6% with a Support Vector Machine. Their standard deviations for the SVM model were also considerably higher (±10.4% compared to 0.54% in our study), though this difference may be explained by measuring different activities and using Leave-One-Subject-Out Cross Validation. Their methodology also indicates that it is better to have the training component in free-living conditions. Other free-living studies have achieved better recognition accuracies but use simple classification problems, for instance Ellis et al. (2014) had an average recognition accuracy of 85.6% but only distinguished sedentary behaviour, standing and walking movement, and driving. Likewise, Fullerton, Heller, and Munoz-Organero (2017) acquired 96.7% accuracy using broad activity label definitions (e.g "Self-Care, "Home Activities"). The recognition accuracies in this investigation were superior to Liu et al. (2011) for a single accelerometer sensor (69.9%), comparable when multiple accelerometer sensors were used (74%), and inferior when they included an additional ventilation sensor (84.7%). This further reinforces that the performance in this investigation could have been improved with the addition of a second ActivPAL on the other thigh.

Generally, there are not many studies which have attempted to differentiate walking across different terrains using only IMU data collected by humans. Hu et al. (2018) appears to be the first study to attempt terrain classification with a single IMU, however this was performed in laboratory conditions and only distinguished between flat and uneven ground. Hashmi et al. (2019) was able to acquire very high classification accuracies (87.5% with an SVM classifier) when distinguishing six types of terrain: concrete, tiles, carpet, asphalt, soil, and grass. Likewise, they also found consolidating terrains into hard and soft categories improved the average recognition accuracy to 92.08%, they however only tested flat terrain and did not include hills or stairs in the classification problem. More recently, Russell et al. (2021) carried out HAR in a cross-country trail. While the study acknowledged the presence of terrain, as well as the speed at which the terrain was traversed, their classification problem was simplified to only distinguish between laying, sitting, walking, running, or climbing a fence. The lack of comparable studies may stem from publication bias: experiments which have attempted similar methodology to this investigation have achieved similar or worse results, and so have elected not to publish results.

The poor cross-validation experiment results in experiment 4 reflect the findings of O'Brien et al. (2017), who tried training data on non-amputated populations and testing them on individuals who had suffered a stroke. In O'Brien et al. (2017), the recall rate of classification accuracy was 53%, increased to 75% when trained on patients with only mild gait impairments. They also had the benefit of a significantly higher training population (15 healthy subjects and 30 stroke patients), so from their results it is inconclusive whether it was the larger training dataset or the condition of the gait that was the key factor in improving performance. Due to only having 4 participants, this investigation did not have the luxury of being able to investigate different types of severity in gait impairment for the amputee population, and all 4 of the participants were considered to only to have mild gait impairment. Therefore, this study cannot implicate how effective the system would be on individuals who for instance have severe comorbidities impacting on their gait mobility. In contrast, Vageskar (2017) was able to achieve much higher recognition accuracies when training on a healthy population, gaining 86.5% accuracy when trained in free-living conditions, however they were able to do so by simplifying the classification problem; most notably, they relabelled stair movement as just “walking” and did not account for sloping movement at all. Including those activities as separate labels would likely have caused a significant drop in their classification performance.

6.4.3 Qualitative Feature Analysis

The handcrafted features that were useful for the supervised investigation were qualitatively analysed through comparisons of the features that are preserved in the two primary dimensionality reduction processes used in this chapter: mRMR feature selection and PCA. It is important to acknowledge that PCA and Filter feature selection, while having a similar goal, achieve the appropriate selection of features through very different approaches. In PCA, the process looks for linear combinations of different features that can effectively explain the total variance of the dataset with compact dimensionality. Whereas a filter feature selection process specifically uses supervision (knowledge of the class) to discriminate which features are more relevant than others. The hypothesis of comparing the two methods was that, if a particular feature demonstrated high discriminative power, and contributed a significant component of the systematic variance, then the feature would be “useful”.

The dataset used in this analysis is the complete dataset for all participants data combined. As this exercise is intended to understand the usefulness of the features across the entire dataset and not to validate the features quantitatively (e.g through accuracy or F1 scores), no arbitrary partitioning is performed on the data. The exercise is conducted at Level 3 of label resolution, though this level of resolution was stated to be untenable from Section 6.4.1, it was still useful using this resolution to show how features are selected from mrMR when a wide pool of activities are available for classification. For the filter approach of this analysis, the mRMR feature selection algorithm is applied, and the scores of each feature are recorded. Features deemed “relevant” were those that proportionally contributed

1% or greater of the total score count across all features, those that contributed less than 1% are also recorded as "irrelevant". To find the relevant features for PCA, PCA is first applied to the dataset and the proportion of contribution to the prinicipal components is plotted, similar to the procedure described in Section 6.2.4 to obtain Fig. 6.5. It was determined that approximately 68 of the first principal components contributed to 95% of the variance. For each principal component, the magnitude of the corresponding values in the eigenvectors of each feature are compared. The features with the highest magnitude indicate they are the primary "contributor" to that principal component (Loukas, 2020). As a result, the features deemed "relevant" by PCA were the unique primary contributors to the first 68 principal components. The "unique" signifies that even if a particular feature was the highest contributor to multiple principal components, it is only counted once. All features that are not the primary contributor to any of the 68 first principal components are recorded as "irrelevant".

The analytical process is applied to the combined dataset of amputees and non-amputated individuals, and additionally applied to separate amputee-only and non-amputee-only datasets. The number of components needed to obtain 95% of the total systematic variance remains the same in all 3 datasets, and so the process of finding relevant features through PCA did not change.

For the combined dataset of all individuals, the features chosen from mRMR and PCA, as well as those that were absent (irrelevant) in either, are represented in the Venn diagram in Fig. 6.12. Generally, it is difficult to comment on which features could truly be considered irrelevant. For example, many of cepstral coefficient features were present in PCA and mRMR, indicating they were relevant, but a handful of the other cepstral coefficients were also absent in both processes. Both feature selection methods contained a combination of statistical, frequency and wavelet features. At most, a handful of wavelet features could have been considered irrelevant; namely wavelet coefficients as formulated by Preece et al. (2009) and Wang et al. (2005), as well as some statistical features including the root mean square, energy, entropy, power, range, mean absolute deviation, and variance. Given the LSTM's comparable performance with the SVM classifier in this chapter using only raw data from the accelerometer, and additionally considering how many components were needed to capture 95% of the systematic variance in PCA, it is reasonable to assess given these results that the handcrafted features chosen for analysis were not well suited to the dataset.

To recap, the features chosen for investigation were derived analytically using a combination of literature and exploratory analysis of the raw data in Chapter 5.7. This process should still be used in future studies as it gives a strong fundamental argument to the feature inclusion, as opposed to blind inclusion. However, while there was thought put into the types of features, the actual features themselves had little thought put into them. Most types of features were applied to all axes of the accelerometer, with some features additionally being applied the derivative of acceleration (the jerk) and the magnitude. Many of these features likely have strong similarities and do not add useful information to the dataset. Likewise, some types of features were very similar to each other. Including the range, interquartile range, as well as maximum and minimum values was likely a redundant approach given their foundational dependence on each other. The wavelet feature sets that have been carried out in previous studies were all functionally similar to one another, with perhaps only one or two sets of these features needed for analysis.

As stated, the analytical process was also applied to separate datasets containing activity data of only the non-amputated individuals, and a dataset containing activity data of only the ILLAs. The resulting Venn diagrams are located in Figs. 6.13-6.14. Despite the reduction in total dataset size, the quantity and number of unique PCA relevant features do not significantly change either in the non-amputated or amputee population models. As PCA is conserving total systematic variance, features that have systematically high variance will generally be conserved regardless of the dataset size. Hence, features such as Cepstral Coefficients, Crest Factor, Skewness, Kurtosis, Spectral Centroids and Median values are present in all three models. On the other hand, mrMR-selected features and subsequently the common overlap in PCA and filter selection processes are greatly reduced with the dataset size. Logically, this appears to be a reasonable response: as the dataset size decreases, there is less intra-activity variation, therefore requiring less features to encapsulate differences between activities. Finally, another Venn diagram was created to look at the overlap of relevant features (combining features conserved in PCA and selected in mrMR) for amputee and non-amputee populations. This Venn diagram is in Fig.6.15. By directly comparing features selected in the amputee and non-amputee population models, it can be seen that while there are different relevant features unique to the amputee and non-amputee population models, no systematic patterns of difference were observed. Each model had a combination of time, frequency and wavelet domain features, and certain features like cepstral coefficients and kurtosis, had unique sub-feature instances in each population model. Since the ILLAs walked over greater varieties of terrains than the non-amputee population, an inference could be made that the relevant features unique to the amputee population model could be important for distinguishing soft terrains and other types of non-concrete hard terrain, but this would require further testing beyond the scope of the thesis.

The primary conclusion from these additional Venn diagrams is that future feature selection processes in these types of investigations should take more time to consider each feature individually and how they corroborate with each other. There also appears to be no systematic differences in feature relevance between ILLA and non-amputee populations, and therefore features chosen specific to the differentiation of activity in a non-amputee population should be translatable to the differentiation of activities in a ILLA population and vice versa.

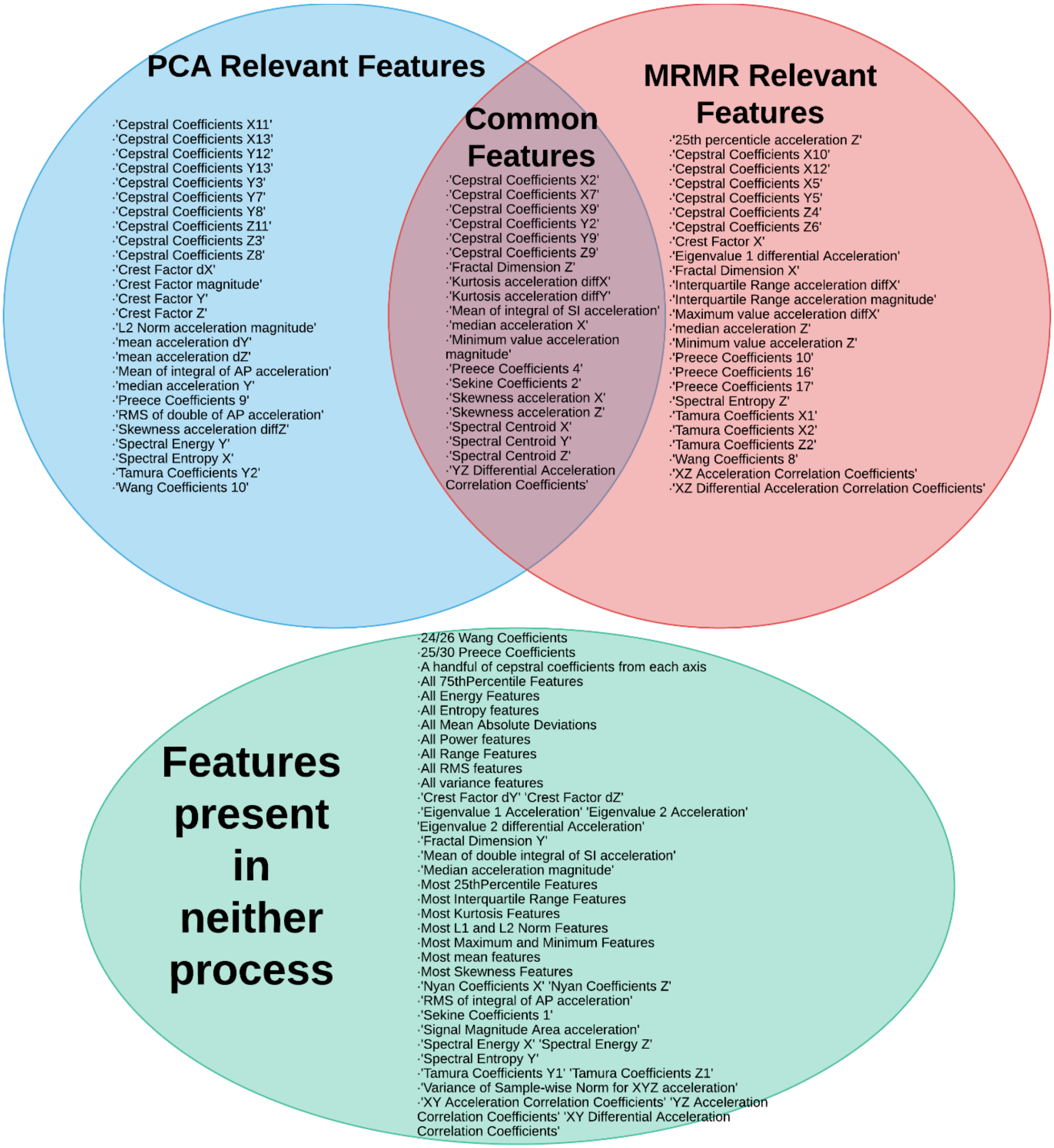


Figure 6.12 Venn diagram of relevant features obtained from analysis of PCA and mrMR, as well as features absent from the relevant list of both processes (Level 3 Terrain Resolution)

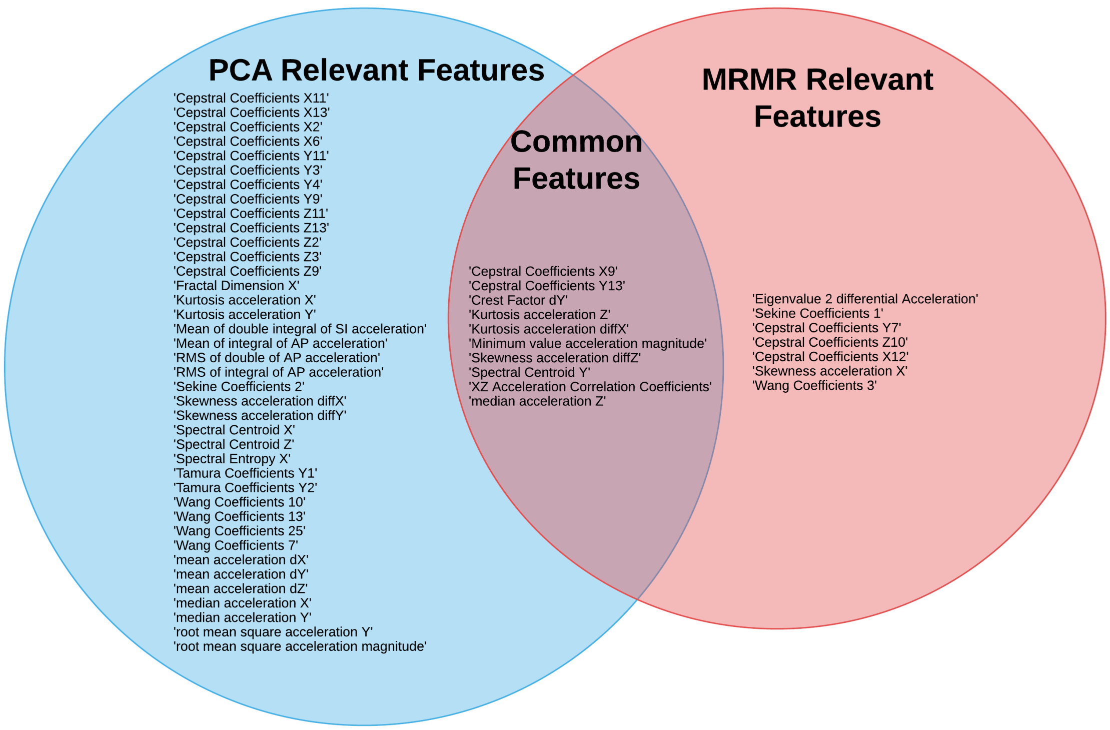


Figure 6.13 Venn diagram of PCA and mrMR relevant features for the non-amputated participants (Level 3 Terrain Resolution)

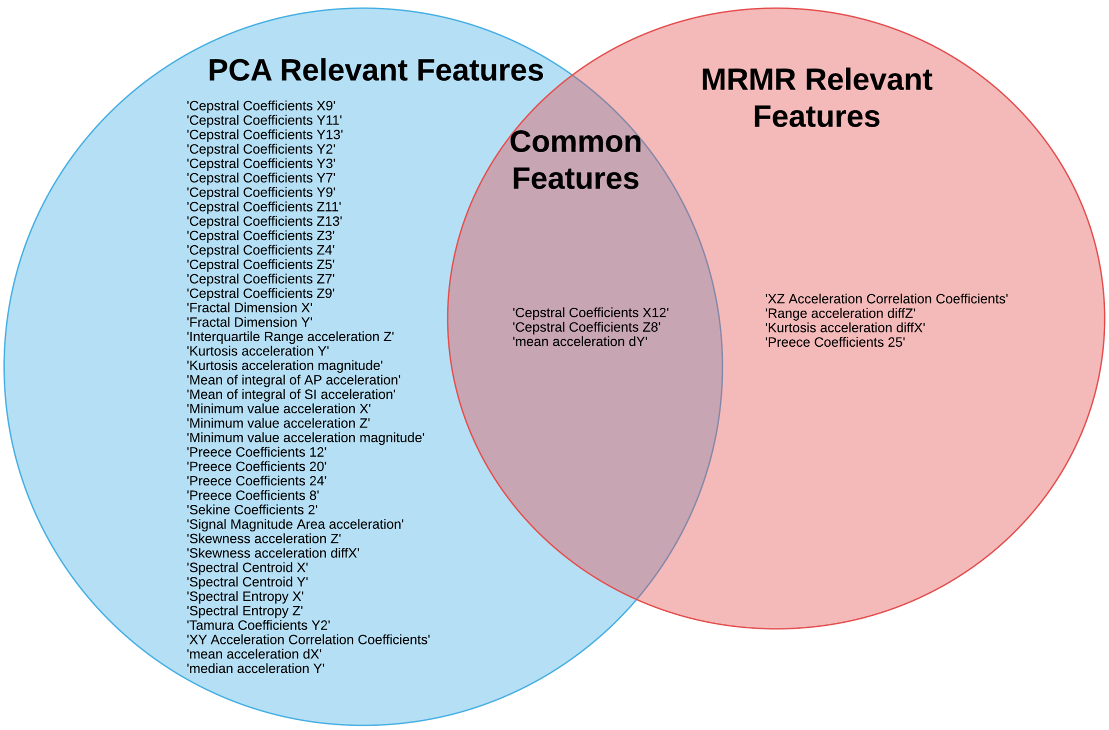


Figure 6.14 Venn diagram of PCA and mrMR relevant features for the ILLAs (Level 3 Terrain Resolution)

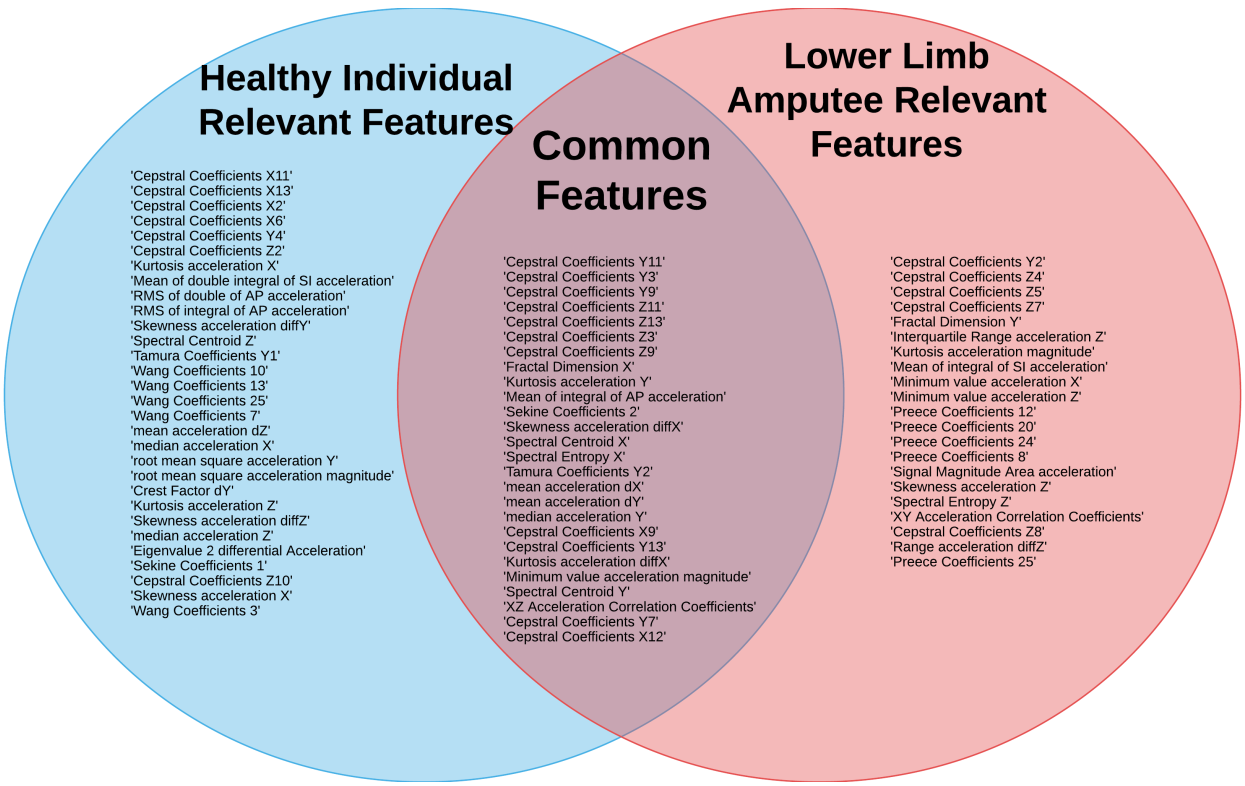


Figure 6.15 Venn diagram of features found relevant for both non-amputated and ILLA demographics. (Level 3 Terrain Resolution)

6.4.4 Misclassification Analysis at the Hard/Soft Terrain Level

To gain a more informed understanding of the classification performance at the hard/soft terrain resolution level (Level 2), which is the highest level of terrain resolution this activity monitoring system could achieve in a realistic manner, an exercise was conducted to identify common "themes" in misclassification via the analysis of the original video recordings. To arrive at this analysis, a holdout cross validation was performed on the combined dataset of ILLAs and non-amputee individuals. Samples in the training partition were trained by an SVM and LSTM classifier, using appropriate hyperparameters defined from experiment 2. The SVM and LSTM classifier were then applied on the testing samples to create independent predictions of physical activities. The resulting predictions were then compared to the ground truth, and subsequently sorted by "strong positive predictions", where both the SVM and LSTM classifier both correctly guessed the activity, and "strong negative predictions", where the SVM and LSTM both guessed incorrectly – and – both classifiers made the same incorrect prediction. For example, if the ground truth label was "Hard, Flat", and the SVM and LSTM both incorrectly guessed "Hard, Uphill", this would be considered a "strong negative prediction". As it would not be practical to manually analyse every single instance of classification or misclassification, five random samples of "strong positive predictions" and "strong negative predictions" for each activity were obtained, resulting in 80 total samples being investigated. Each sample had a corresponding timestamp which allowed the (mis)classified activity to be reviewed in the video footage. The key summary of the analysis is hereby detailed:

When analysing the "strong positive classifications", there were some interesting observations made with the minority classes. Hills of either hard or soft terrain appeared to have strong classifications results when the measured hill angles were very steep. Logically, this makes sense; a steep slope will have a greater impact on the subject's gait, which should reflect in changes to the feature vectors of these samples. Correctly classified stair movement happened typically with narrow flights of stairs with high stair heights. Again, these types of stairs would have a large impact on gait trajectory which should reflect in differences in the feature vectors. Correctly classified soft terrains would typically have physical properties that would increase the differences in gait versus a hard terrain. For example, correctly classified grassy terrain (considered “soft”) would typically feature long cuts of grass and leaves lying on the ground, both of which can act to reduce the GRF during gait. In sandy terrain (another terrain considered “soft”), correctly classified instances would demonstrate very soft sand: in the videos, there would be deep footprints in the sand from people previously walking along the path. Moreover, it was clearly visible in these correctly classified sandy segments that the entire ground would deform and stretch over the foot during gait. These would reduce the GRF experienced by the subject, thereby creating a change in the responsive acceleration behaviour.

In contrast, soft terrains that were strongly negatively misclassified exhibited properties that made them more closely resemble harder terrains. Misclassified grassy terrain would exhibit short grass on what appeared to be hardened soil, while misclassified sandy terrain would appear as hard-packed sand with barely visible footprints. The other strongly misclassified activities revealed interesting observations. Stair misclassification appeared to mostly arise from when subjects traversed landings that are interspersed between flights of stairs. During the annotation process, these flights of stairs would be “ignored” (meaning still considered downstairs) if the landing was short enough (roughly less than a full stride long). However, this analysis indicates that even these short landings can have an impact on misclassification and should be discarded from training data in future studies. Misclassification of flat and hilly terrain mainly arose as a result of imprecision of elevation readings from the Strava application. If the readings were infrequent (between five and fifteen seconds), this would not build an accurate estimate of the elevation data, in turn impacting on the calculated hill angle and the subsequent annotation. Upon reviewing the misclassified hills traversed in Google Maps, the misclassification coming from the classifier could be interpreted as the absolute correct activity, while the annotated activity could in fact be incorrect. There were a not-insignificant number of instances where “soft, downhill” was misclassified as “downstairs”. This was caused by participants walking down the (grassy) hill at a 45-degree angle, always leading with one foot in front of other during the descent (as if climbing down a steep slope), this deviation in movement from gait that would be exhibited while climbing down a gentle slope likely caused the misattribute of the label. This behaviour was not exhibited in hard, downhill classifications, likely due to higher friction on the surface giving subjects a greater grip and easing their confidence at descending the slope. Moreover, there were several instances of “Hard, Uphill” misclassified as “Soft, Flat”. This confusion may originate from increased stance duration and step time on inclined slopes (Han et al., 2009), which may be equivalized in the feature matrix to slower gait movement on soft terrain where the ground is less stable. For example, in sandy terrains, the foot sinks into the sand on impact, thereby requiring more time to lift the foot off again, resulting in increased stance time.

The key findings of this analysis were the discovery of “intra-terrain” variation and errors in the annotation process. The intra-terrain variation was discovered by observing that some terrains exhibited different properties that could construe them either as hard or soft terrain. It is not possible to validate a terrain’s hardness or softness purely through visual inspection, so perhaps a more constructive solution is to develop a multi-terrain laboratory environment where the hardness of terrain is known and can be directly controlled. This would of course limit the generalizability of findings to real outdoor conditions, but the trade-off would be the ability to train the system with data labels that are closer to the “absolute truth”. Errors in the annotation process meanwhile indicated that the predicted labels from the classifiers were on occasion, more accurate to the absolute truth, which would imply that the classifiers had greater performance than suggested in the results. More research is required for providing a better estimation of the absolute truth, particularly in the identification of hill segments. As discussed in Chapter 5.5, only one round of annotation was performed by a single author, which could have easily led to some mistakes in the annotation process - particularly in cases of terrain transitions. A basic solution to improve accuracy of the GTA would be to repeat rounds of annotation with other researchers and obtaining an "average" label. Other solutions could include using a second mobile device with Strava, leaving the final elevation reading to be an average of the two readings, or to have the participant signal to the camera when they are experiencing uphill or downhill movement. Such readings would be prone to subjective errors from the participant but would provide an additional point of unreliable information. Finally, the limitations of this analysis should be considered. This experiment only looked at random instances of misclassification but would probably be more informative if it had looked at all possible combinations of misclassification (for example, comparing a misclassified "Upstairs" to "Downstairs", "Downhill", "Uphill" etc.).

6.5 Chapter Conclusion

This chapter was an investigation of how supervised learning systems can recognize activities carried out by non-amputee individuals and those with lower limb amputation, finding good results but with room for improvement. Generally, there was scope to expand walking activities to include hard and soft terrains alongside walking uphill, downhill, upstairs and downstairs, and analysis of misclassified activities revealed some key weaknesses in the methodology, further indicating the systems performed better than presented in the results. It was not feasible to try and recognize ILLA activities when the system was trained only on non-amputee individuals or other ILLAs, which was most likely due to the large class imbalances. The analysis of relevant features and misclassification instances have constructively identified the flaws in the feature engineering approaches and annotation processes that were carried out in this research, and can be improved upon in future studies. The next chapter will look at the same dataset, this time taking an unsupervised approach to try and build a cluster analysis of these activities without the use of annotated labels.

Chapter Seven

Unsupervised Cluster Analysis of Walking Activity Data for Non-Amputated Individuals and Individuals with Lower Limb Amputation

7.1 Introduction

This chapter is an exploration of the application of unsupervised learning and cluster analysis for human activity recognition, including an ILLA population. An unsupervised approach differs from a supervised approach in that knowledge of the label is not used to train the machine learning algorithm. Instead, the general objective is to present the data in a visualizable format (between 1 and 3 dimensions) where datapoints with similarities converge together into clusters. The grouping of data into clusters can indicate unforseen similarities or differences in the data that was not expected prior to conducting the investigation. Given the inherent disadvantage of being unable to use the Ground Truth to influence classifier boundaries, unsupervised approaches will generally underperform compared to supervised approaches when applied in HAR (Attal et al., 2015; Trabelsi et al., 2013). However, from a clinical perspective, it is still worth exploring unsupervised approaches. In a scenario where a clinical researcher wants to monitor activity of their patient, they may not have the financial resources to acquire a trained classifier, or the skill to develop their own. From that perspective, an unsupervised model would be able to interpret useful data in a cost-inexpensive fashion, only requiring that the client should wear an activity monitoring device for a brief period of time. Additionally, as was seen in Chapter 6.3.4, it was indicated that separate trained supervised classifiers may be required for ILLAs with different conditions of amputation, and so collection of data for trained classifiers for different types of amputees could become overwhelmingly expensive and difficult to obtain. This investigation is a proof of concept that clinically relevant data of a lower limb amputee can be obtained without the use of a trained supervised system.

For a general population, Ariza Colpas et al. (2020) provides a concise summary of the work done for unsupervised approaches to HAR. Unsupervised models that use wearable sensors as primary means of data acquisition are primarily used for the recognition of low-level activities such as standing, sitting and lying (Al-Ani, Le Ba, and Monacelli, 2007; Bao and Intille, 2004; Mathie et al., 2003), though a few studies have attempted to distinguish higher-level sporting activities and activities of daily living (Chambers et al., 2002; Huynh, 2008; Lester, Choudhury, and Borriello, 2006). As previously stated, this is first study to attempt to cluster physical activity within an ILLA population (Labarrière et al., 2020). Being unable to monitor indoor activity due to ethical restrictions, the investigation focuses on low-level walking activities that are carried out in outdoor environments.

The first part of the methodology in this chapter will document the journey of investigating at the data collected from Chapter 5 from an unsupervised perspective and details how an experiment with viable results was constructed. The second part will then go into details about the experiment itself and the associated outcomes.

7.2 Methodology

7.2.1 Pre-Processing

The data preparation steps taken in this investigation are identical to those carried out in Chapter 6 and explained in detail in Chapter 5.

7.2.2 Part I: Obtaining Viable Clustering Models

There were two main objectives to meet in the unsupervised clustering of the feature data:

A) Obtain a meaningful, visualizable cluster model.

B) Construct a cluster algorithm that can robustly highlight important clusters in the data.

Objective A had to be dealt with before any meaningful work on objective B could begin. For the investigation, the general analytical approach to achieving objective A was to look at how the activities were clustered in a low dimensional space (providing their ground truth labels for illustration) and then systematically work towards a clustering model in which activities were distinctively separated into different clusters. In this investigation, one of the main challenges of working towards achieving objective A was considering the factors involved in the construction of the low dimensional cluster model, and the order in which the factors should be approached. The main factors to consider were:

- **Model Population:** Should the data for constructing a cluster model include all participants, split between non-amputated and ILLA demographics, or create models for each subject individually?
- **Dimensionality Reduction:** Which dimensionality reduction method is the most appropriate?

- **Parameter Tuning:** What are suitable hyperparameters for the dimensionality reduction method?

- **Label Resolution:** How far does the detail of the true labels need to be reduced before patterns in the clustering are observed?

It was impractical to try and explore all possible combinations of the above parameters, as analysis at this stage is primarily performed on a subjective basis by looking at the resulting clustering models. Instead, a series of small analytical exercises were carried out:

1. Using Level 1 of Label Resolution from 6.2.3 ("no terrain resolution") and any viable dimensionality reduction method with default parameters where applicable, decide on the appropriate model population.

2. Using the model population from exercise #1 and Level 1 of label resolution, decide on the best dimensionality reduction method, again using default parameters where possible.

3. Using results from exercises #1 and #2, decide on appropriate level of label resolution

4. Tune the dimensionality reduction method for its most suitable parameters

7.2.3 Exercise #1: Model Population

The first stage was to determine whether the model should include all participants, use separate models for each individual, or to split models into ILLA and non-amputated populations. Analysis was performed with 3 dimensionality reduction techniques, namely, PCA, tSNE, and UMAP. The processes behind these techniques were explained in Chapter 4, Sections 4.2.5.2.1 & 4.2.5.2.2. An additional non-linear dimensionality reduction technique previously described, kPCA, was not used in this step, this is because the kernel procedures in kPCA require a creation of a kernel matrix ($n_s * n_s$), where n_s is the number of samples.

While this method is viable for looking at individual models, the required matrix size for ILLA/non-amputated models and all-subjects-combined model exceeded the RAM capacity for the testing hardware. For each dimensionality reduction method, the dimensionality is reduced from the original dimensionality (243) to 2 dimensions. In PCA and kPCA, this is done by calculating the $n_f * n_f$ coefficient matrix (n_f being the number of features), conserving the first two columns of coefficients, and multiplying the original dataset by the coefficients. The data points are plotted in a 2D space and are labelled with their base activity label (flat/level walking, uphill, downhill, upstairs or downstairs) to indicate the effectiveness of the method. The outcomes of this exercise are shown in the series of Figs. 7.1-7.12. For brevity, only one subject from the Individual Model approach is illustrated (subject H1).

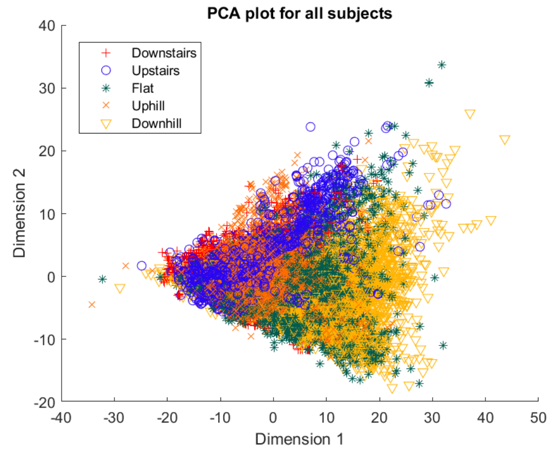


Figure 7.1 Activity data reduced to 2D via PCA, for all participants

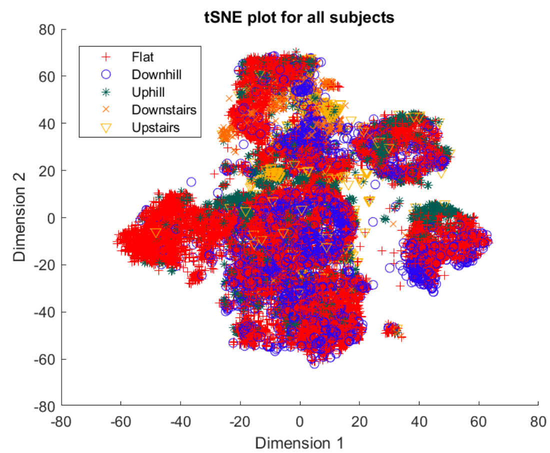


Figure 7.2 Activity data reduced to 2D via tSNE, for all participants

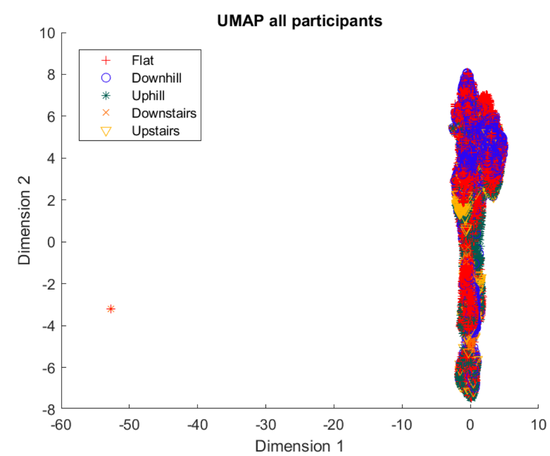


Figure 7.3 Activity data reduced to 2D via UMAP, for all participants

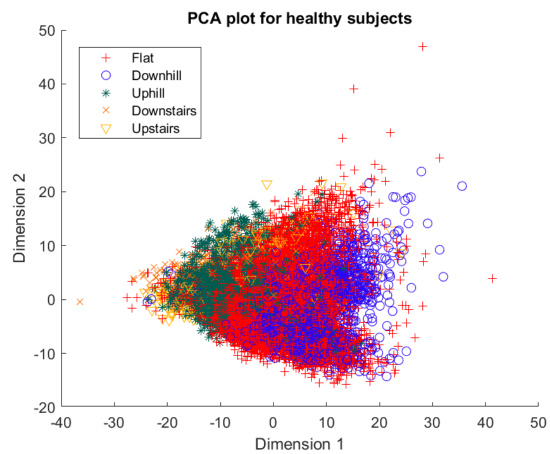


Figure 7.4 Activity data reduced to 2D via PCA, for non-amputated individuals

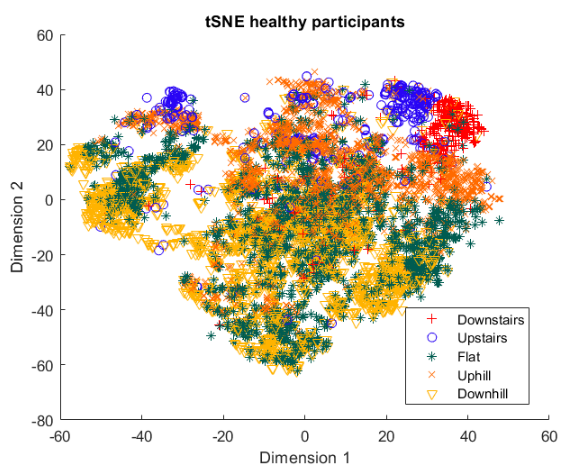


Figure 7.5 Activity data reduced to 2D via tSNE, for non-amputated individuals

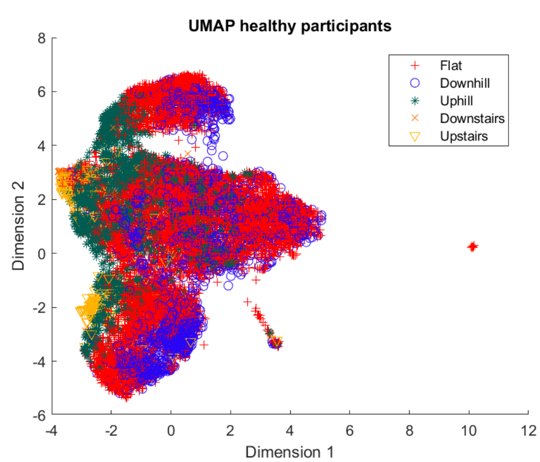


Figure 7.6 Activity data reduced to 2D via UMAP, for non-amputated individuals

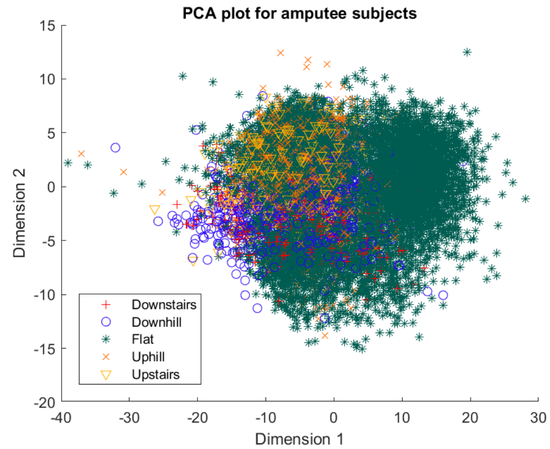


Figure 7.7 Activity data reduced to 2D via PCA, for ILLAs

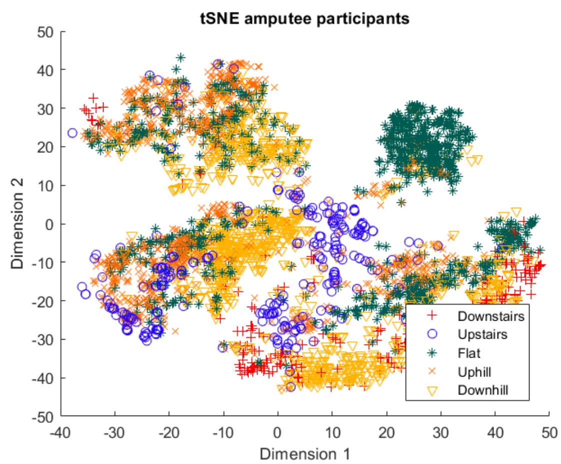


Figure 7.8 Activity data reduced to 2D via tSNE, for ILLAs

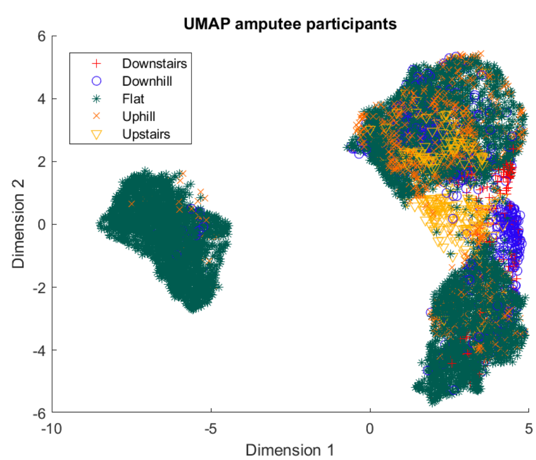


Figure 7.9 Activity data reduced to 2D via UMAP, for ILLAs

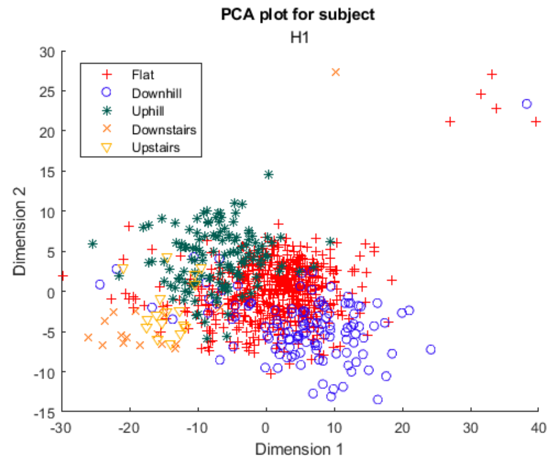


Figure 7.10 Activity data reduced to 2D via PCA, for participant H1

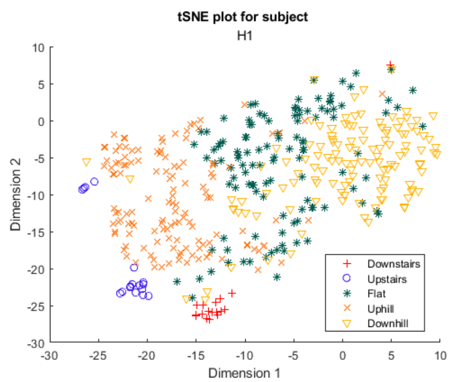


Figure 7.11 Activity data reduced to 2D via tSNE, for participant H1

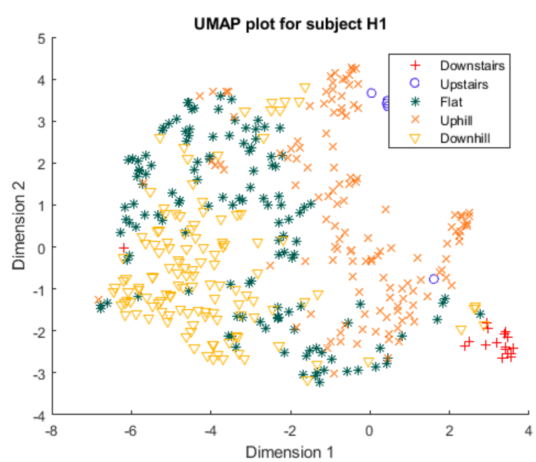


Figure 7.12 Activity data reduced to 2D via UMAP, for participant H1

From Figs. 7.1-7.12, all three dimensionality reduction approaches are only viable when cluster models are created for the individual subject. In tSNE and UMAP models, when the activity labels are replaced with the identity of the subject, as seen in Fig. 7.13, it becomes clear that these approaches have modelled their cluster formation on the subject, rather than the activity. While there is a slight gradient-like distribution of the activity classes in the PCA models when viewing the all-subjects-combined and non-amputated-subject models (Figs. 7.1 & 7.4, there are no visually distinct cluster formations. There are also no appreciable formations of activity classes in the ILLA only model. Thus, the chosen approach was to use individual models for cluster analysis.

7.2.3.1 A note on class imbalance

In the supervised chapter, the imbalance between classes was dealt with by applying SMOTE to prevent the classifiers from overfitting on the majority class. Balancing the dataset requires knowledge of the label, which is inherently a supervised technique. Additionally, oversampling via SMOTE has a strong influence on the mapping behaviour, particularly in non-linear dimensionality reduction techniques. The mapping techniques are able to recognize the low variance in the synthetic minority data samples that are generated through SMOTE, and will map them closely together in lower dimensions, which can result in clustering models such as Fig. 7.14. Here, the minority classes are the upstairs and downstairs labels, which have resulted in an orthogonal-shaped clustering model. While this shows good separability of the stair classes, the model shape could change entirely depending on the ratio of the minority classes to the majority classes. Thus, the datasets for each individual are not balanced, to try and give a realistic representation of how the clustering algorithms would perform on a true unlabelled dataset.

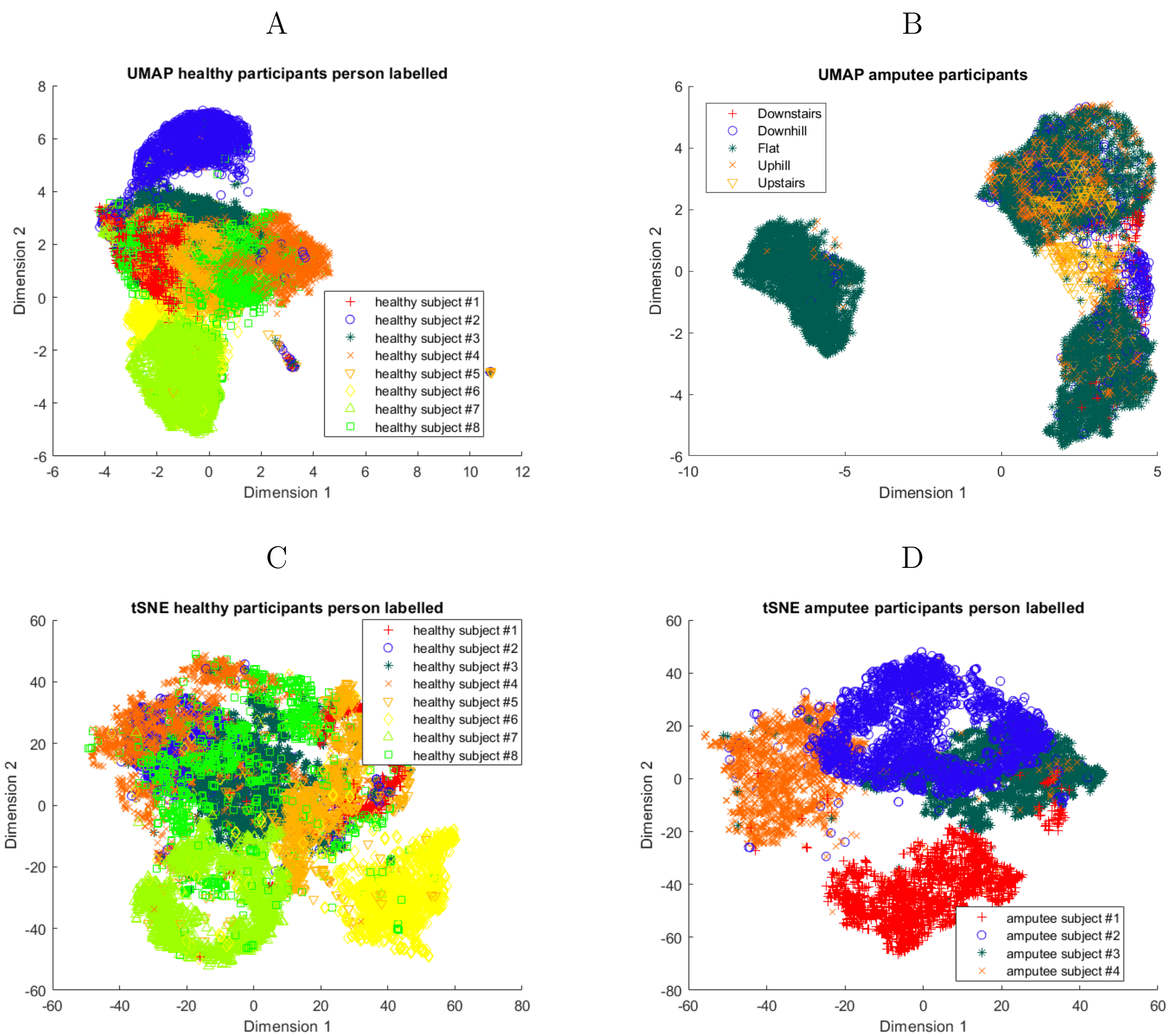


Figure 7.13 A: Data of non-amputated participants after UMAP dimensionality reduction, labelled by the contribution of each participant. || B: Data of ILLA participants after UMAP dimensionality reduction, labelled by the contribution of each participant. || C: Data of non-amputated participants after tSNE dimensionality reduction, labelled by the contribution of each participant. || D: Data of ILLA participants after tSNE dimensionality reduction, labelled by the contribution of each participant.

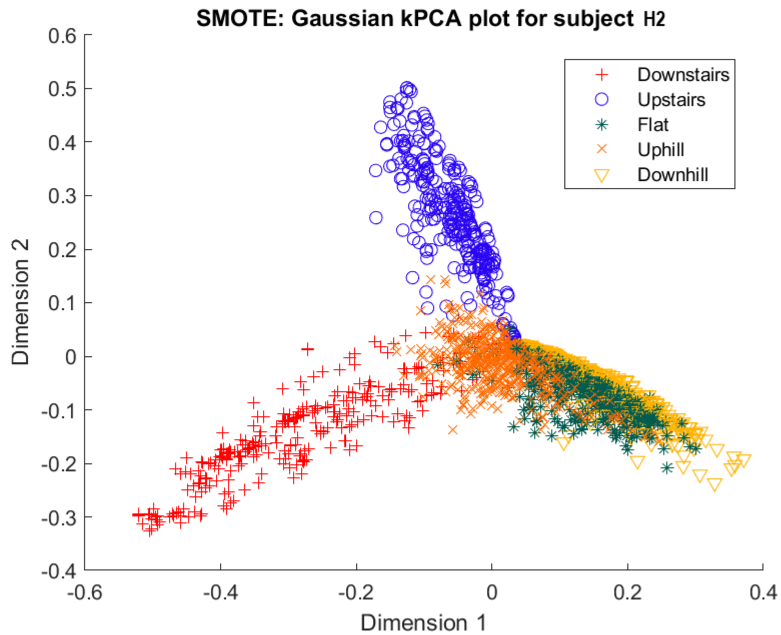


Figure 7.14 Cluster model of participant H2 after Gaussian kPCA dimensionality reduction

7.2.4 Dimensionality Reduction Method

With the appropriate population model chosen, the next step was to select which of the dimensionality reduction techniques were most suited to the investigation. Alongside PCA, tSNE and UMAP, kPCA models with various kernels were included in this exercise. The kPCA models were manually tuned to have sufficient kernel widths by visual inspection of one non-amputated and one ILLA subject, then expanding to include all individual models to ensure the parameter was universally appropriate. Once again, all dimensionality reduction methods were plotted in a 2D space and were colour-coded by their true activity labels, which can be seen in Figs. 7.15-7.28. For sake of brevity, not all images from each participant are shown. Instead, each dimensionality reduction method shows a cluster model from one non-amputated individual (participant H6) and one ILLA (participant A3).

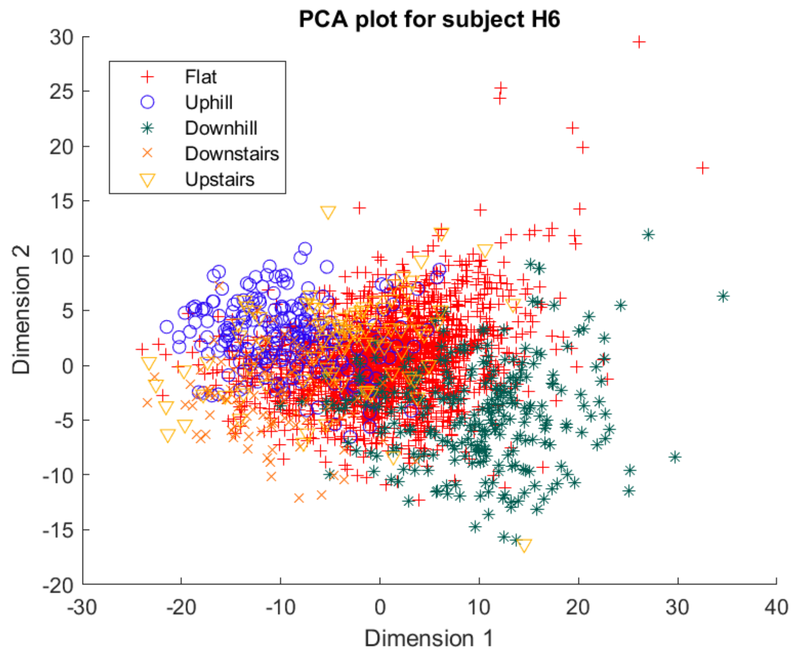


Figure 7.15 PCA-created cluster model on a non-amputated individual

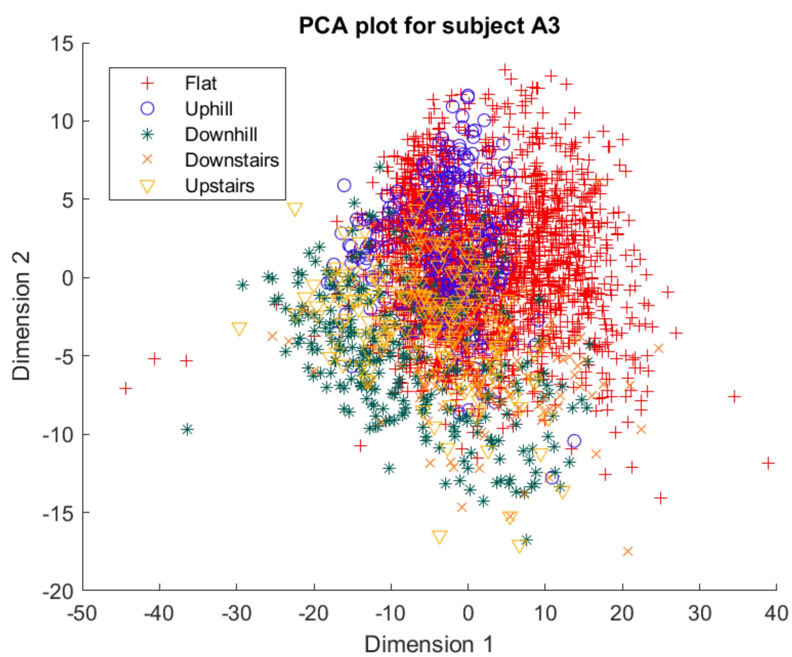


Figure 7.16 PCA-created cluster model on an amputated individual

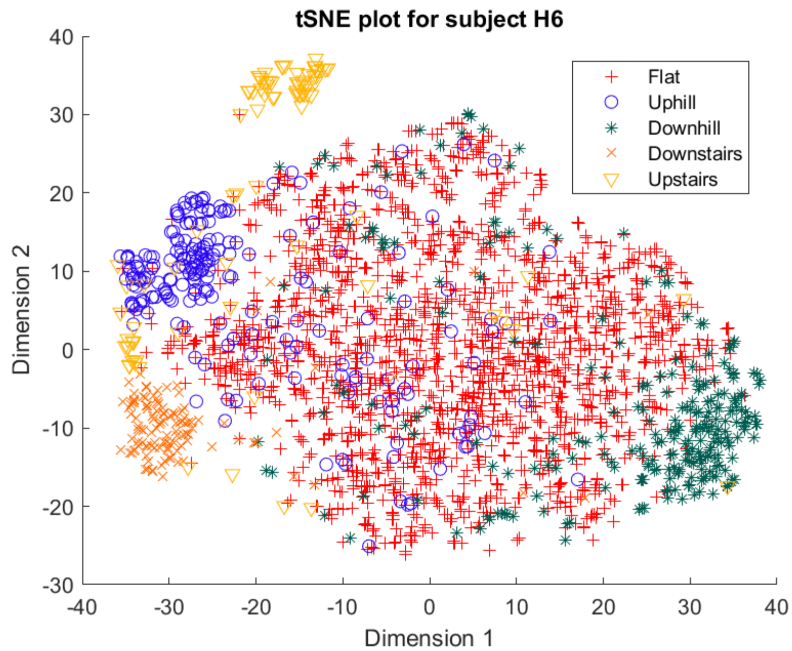


Figure 7.17 tSNE-created cluster model on a non-amputated individual

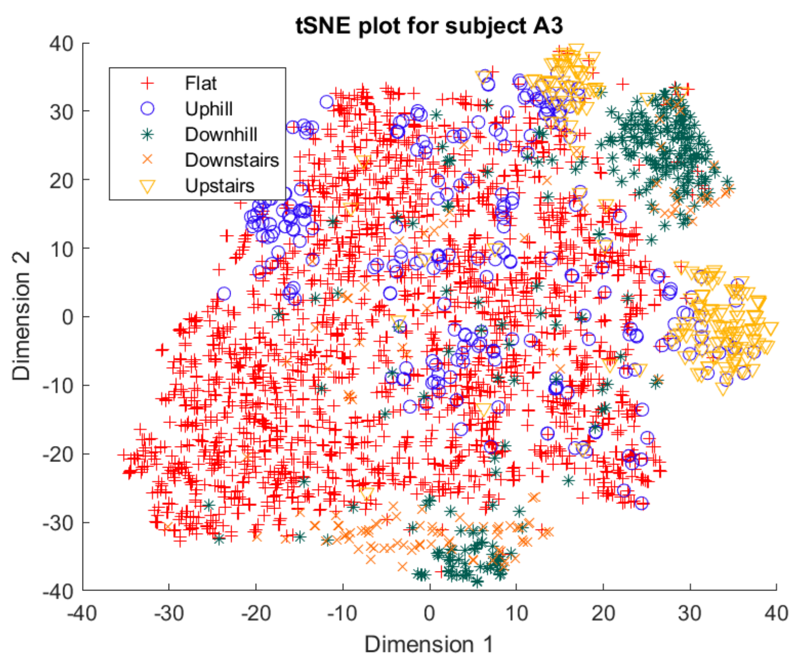


Figure 7.18 tSNE-created cluster model on an amputated individual

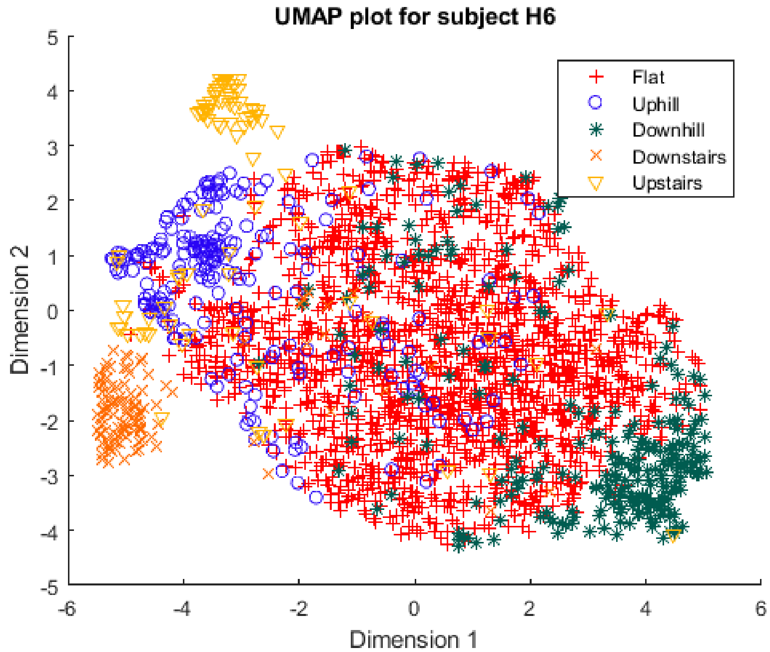


Figure 7.19 UMAP-created cluster model on a non-amputated individual

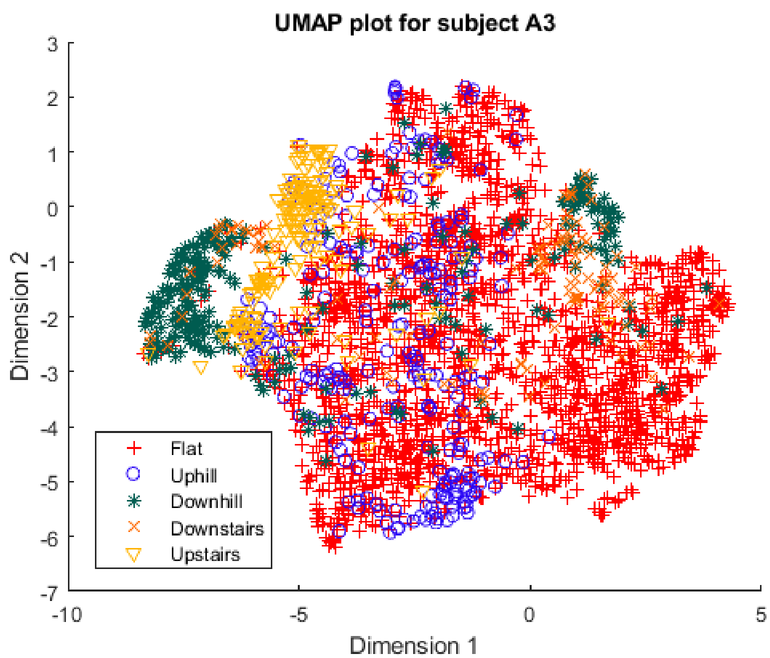


Figure 7.20 UMAP-created cluster model on an amputated individual

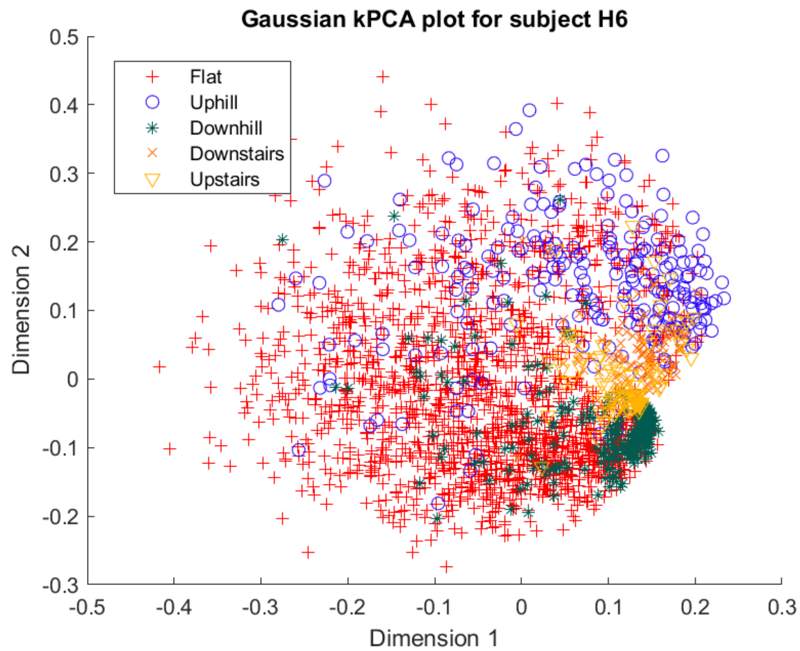


Figure 7.21 kPCA (Gaussian Kernel) created cluster model on a non-amputated individual

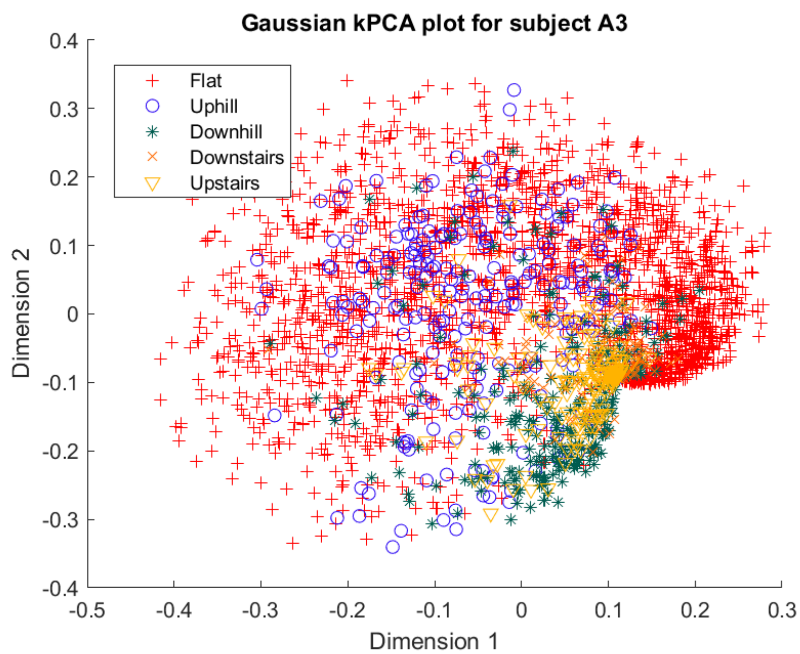


Figure 7.22 kPCA (Gaussian Kernel) created cluster model on an amputated individual

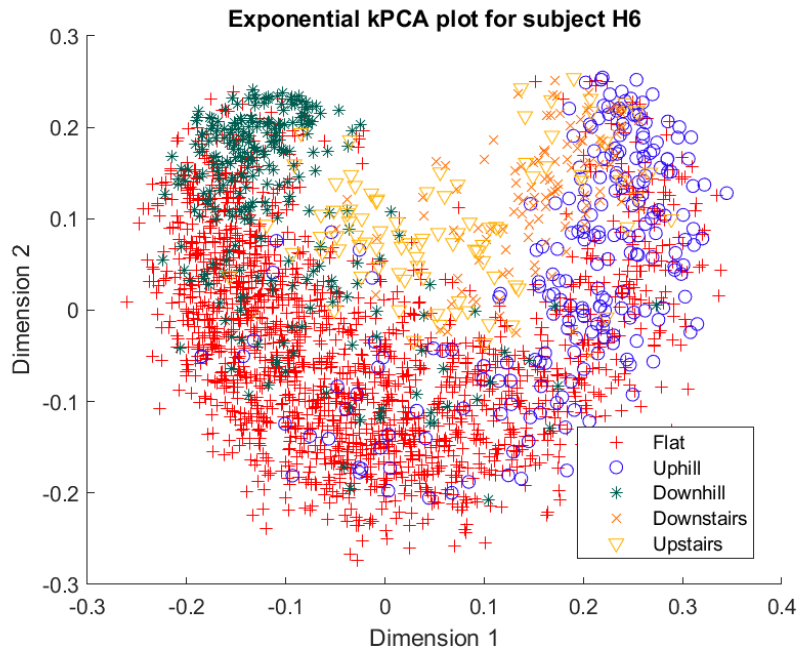


Figure 7.23 kPCA (Exponential Kernel) created cluster model on a non-amputated individual

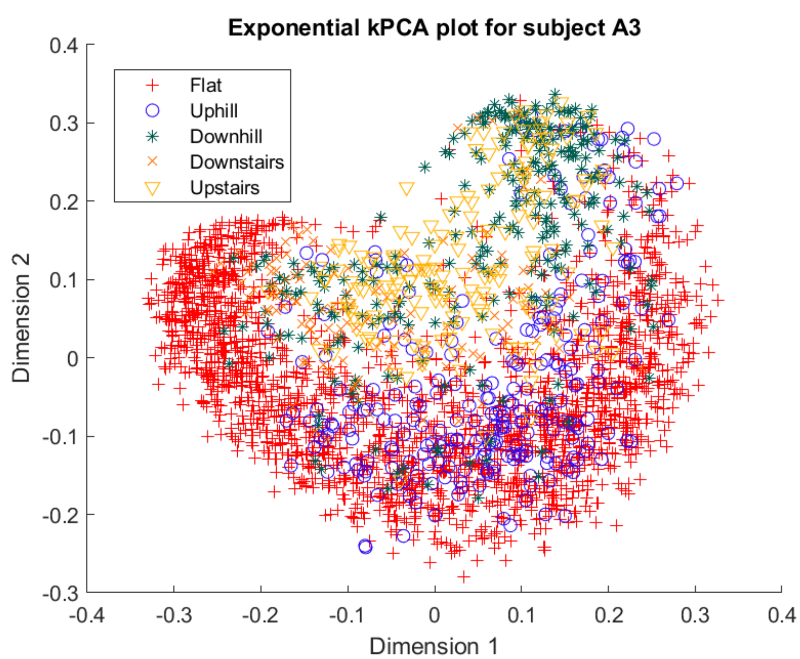


Figure 7.24 kPCA (Exponential Kernel) created cluster model on an amputated individual

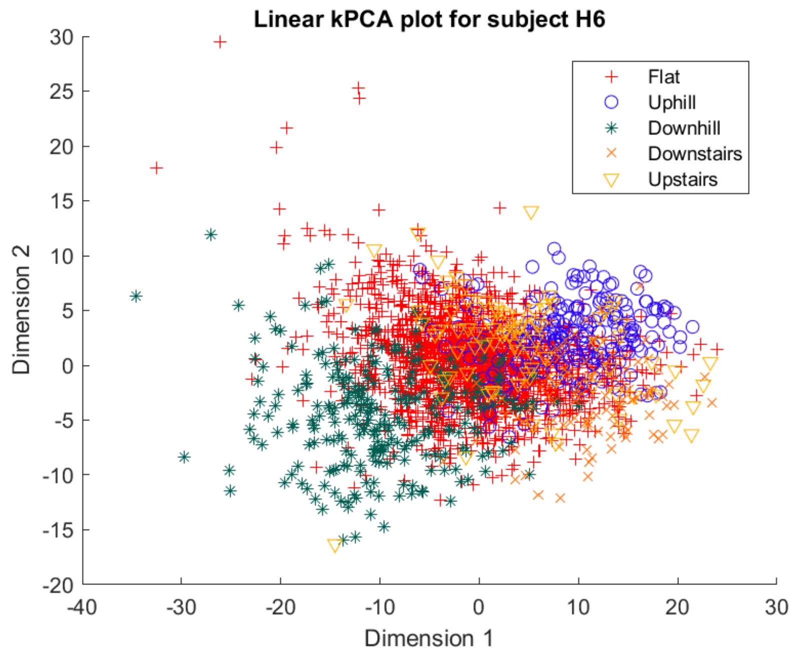


Figure 7.25 kPCA (Linear Kernel) created cluster model on a non-amputated individual

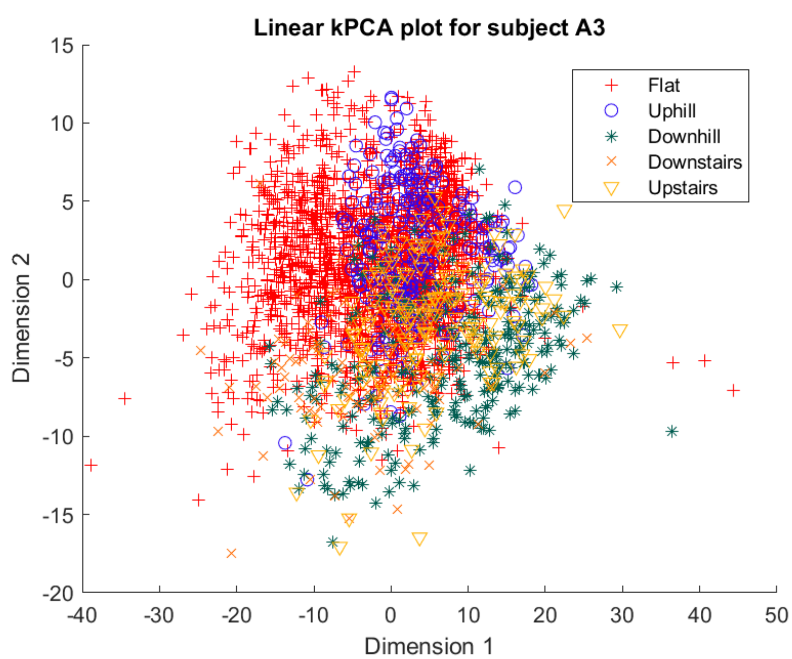


Figure 7.26 kPCA (Linear Kernel) created cluster model on an amputated individual

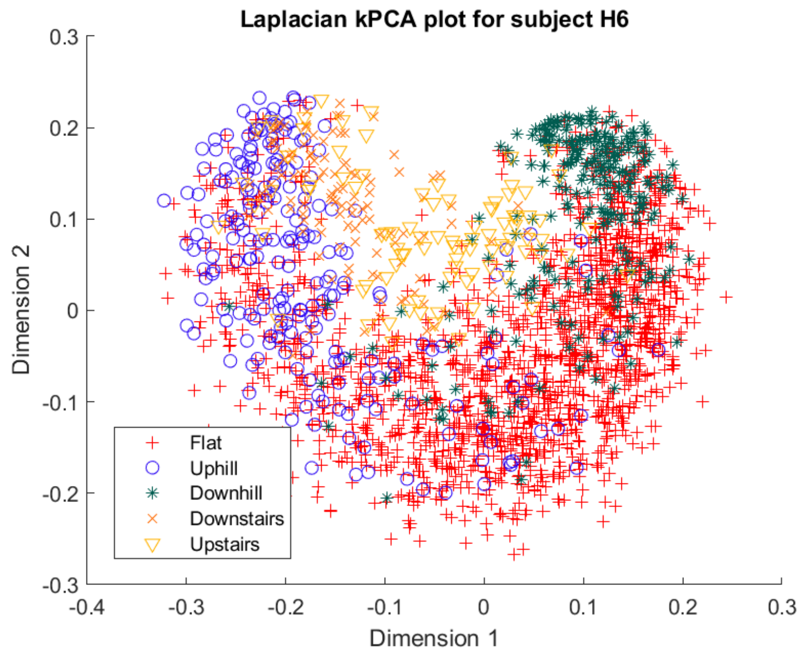


Figure 7.27 kPCA (Laplacian Kernel) created cluster model on a non-amputated individual

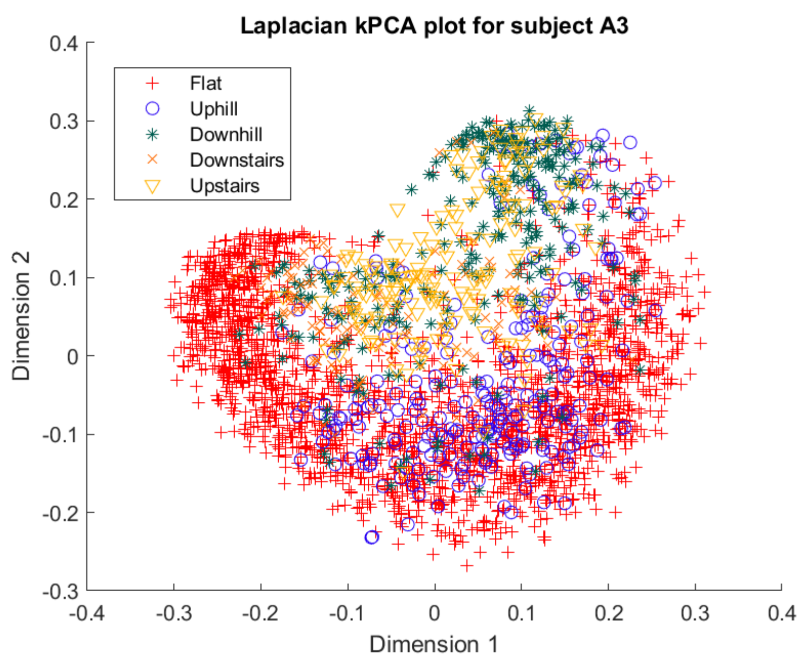


Figure 7.28 kPCA (Laplacian Kernel) created cluster model on an amputated individual

Figs. 7.15-7.28. revealed that the tSNE and UMAP models were the only models to show appreciable separability of the activity classes. In PCA (Figs. 7.15, 7.16) and all kPCA models (Figs. 7.21-7.28), there was complete failure to distinguish any of the 5 classes, and all data points had merged into a singular homogenous shape. Recalling back to Fig. 6.5 in Chapter 6, it was discovered that it took around 70 components to explain 95% of the total systematic variance in the combined dataset of non-amputated individuals and ILLAs. This finding remains true when PCA is applied to each individual's dataset separately. Therefore, it is unsurprising that the PCA process is unable to properly distinguish activities with just 2 principal components. Likewise, the relevance of the principal components in kPCA can be determined by examining the ratio of the sum of the conserved eigenvalues over the total sum of all eigenvalues. Taking Gaussian kPCA as an example, these ratios ranged from 0.023 to 0.029, showing the first principal components of kPCA also comprise very little of the total systematic variance.

The original intention from this exercise was to proceed using both tSNE and UMAP, as they had very comparable performances on a per-subject basis. However, there were some issues with community written UMAP implementation for Matlab (Meehan et al., n.d.). Due to an unexplained instability in the UMAP function, the Matlab environment would consistently crash whenever the algorithm ran over many iterations. As a result, it was decided to only use the more stable tSNE function for the remainder of the investigation.

7.2.5 Exercise #3: Selecting Appropriate Level of Terrain Resolution

Following exercise #2, the tSNE models for each subject were analysed for various levels of terrain resolution, refer to Chapter 6.2.3 for an explanation of each level. When the terrain resolution is increased to Level 2 (including hard and soft terrains), the non-amputated individual models are largely unaffected due to almost exclusively walking on hard terrain. However, it became immediately apparent when viewing the ILLA cluster models that the dimensional mapping was unable to differentiate hard and soft terrains. It was thus seen as redundant to try and increase the label resolution any further to Levels 3, 4 or 5, which include even more labels (or are based only terrain, which is largely irrelevant to the main investigation). Even at Level 1 of terrain resolution, there were some concerns that the models were unable to separate flat, uphill, and downhill data. Thus, an additional simplified level of resolution was introduced, termed Level "0". In this level of resolution, the flat, uphill, and downhill labels are all consolidated into a single label – "Walk". The results for non-amputated and ILLA subjects are illustrated in Figs. 7.29-7.34.

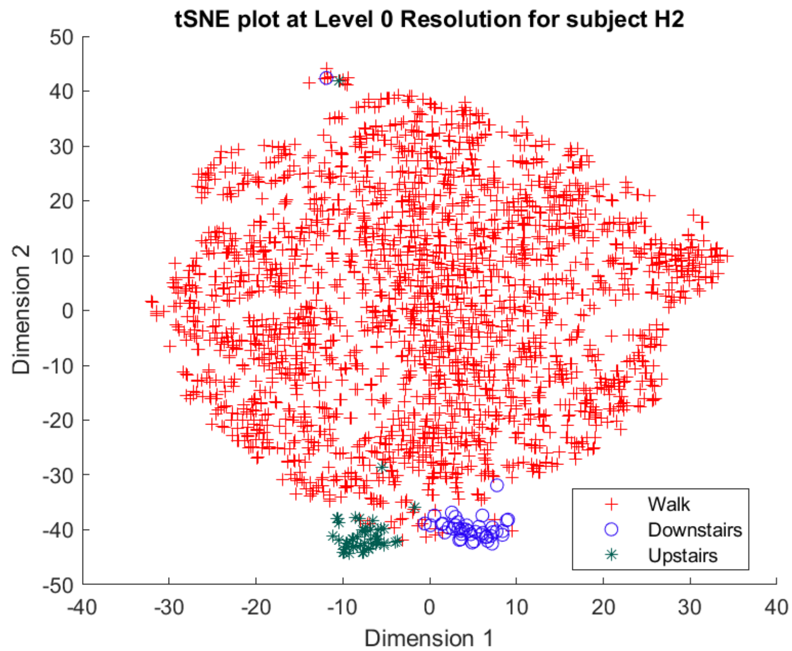


Figure 7.29 tSNE-created cluster plot of non amputated individuals for terrain resolution Level 0

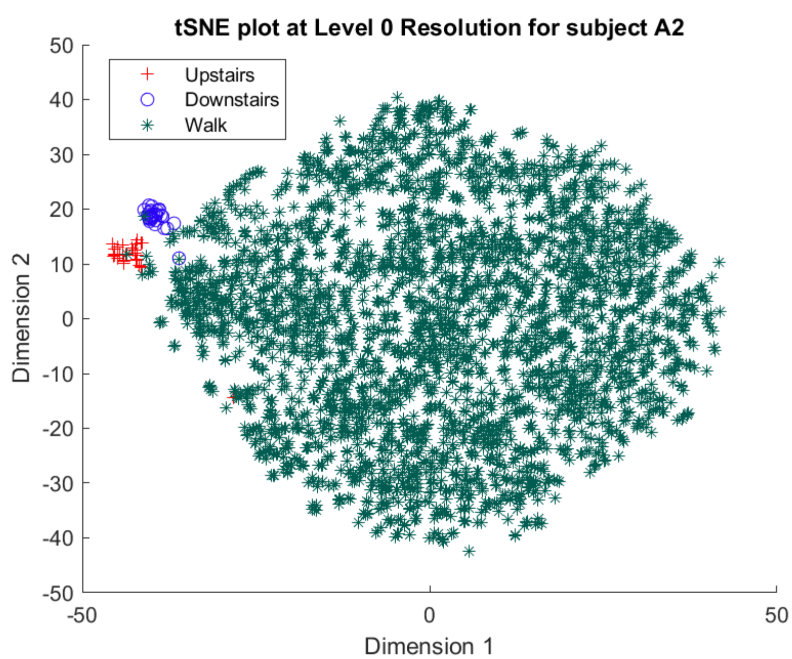


Figure 7.30 tSNE-created cluster plot of amputated individuals for terrain resolution Level 0

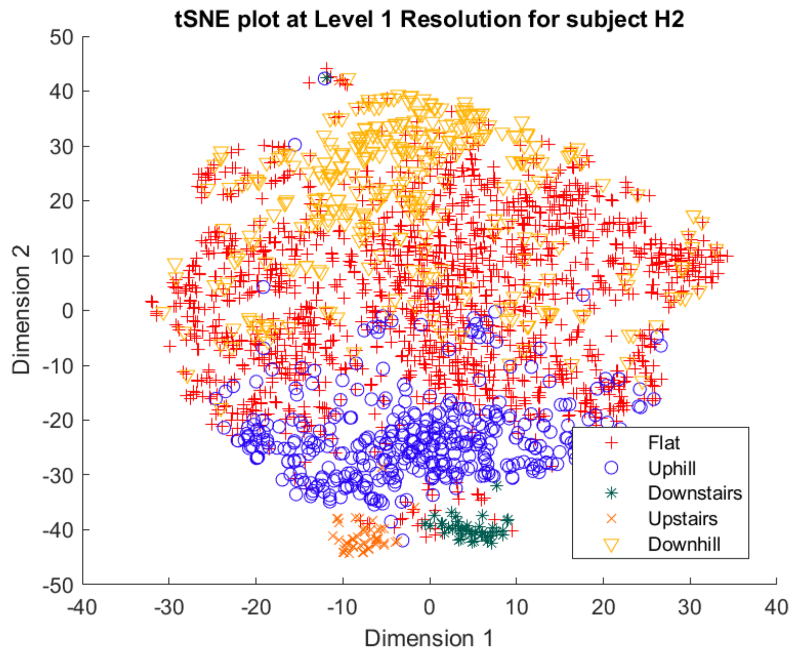


Figure 7.31 tSNE-created cluster plot of non amputated individuals for terrain resolution Level 1

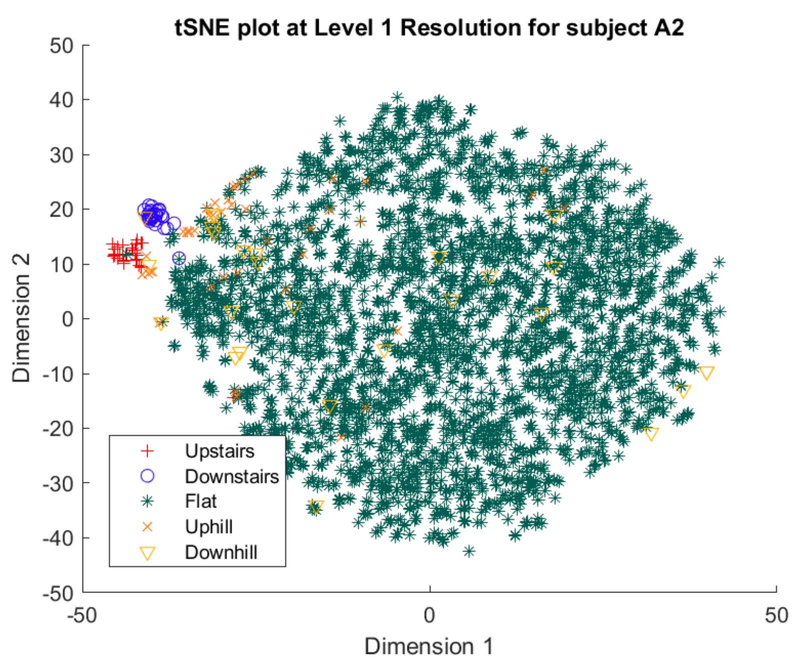


Figure 7.32 tSNE-created cluster plot of amputated individuals for terrain resolution Level 1

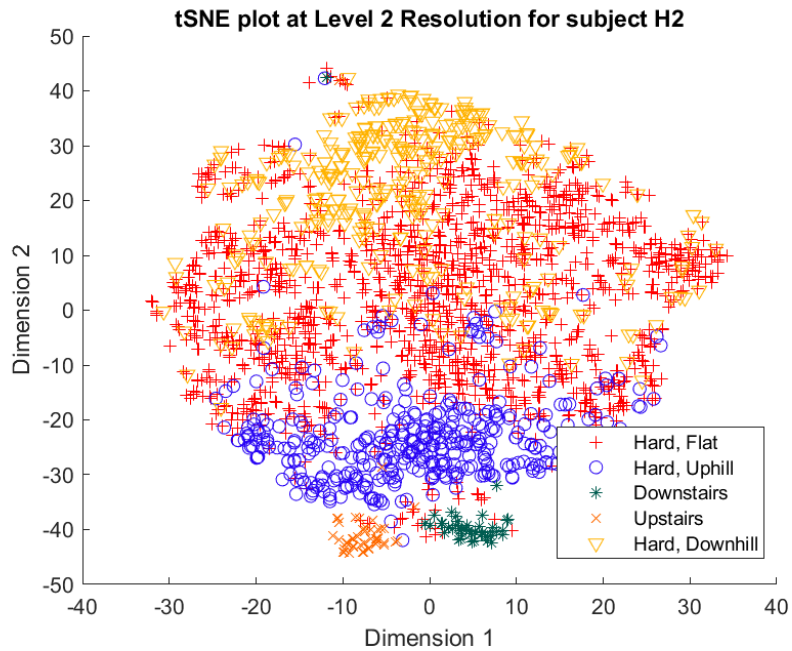


Figure 7.33 tSNE-created cluster plot of non amputated individuals for terrain resolution Level 2

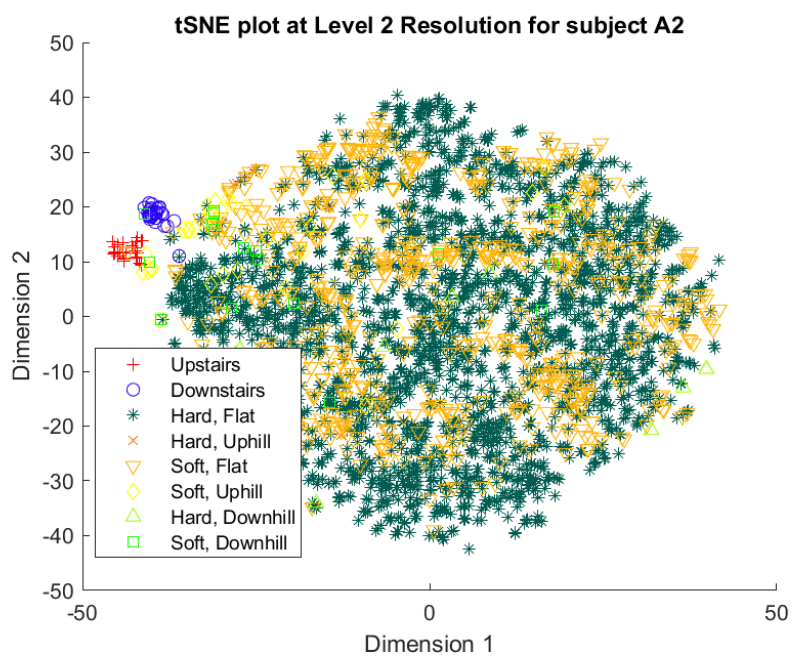
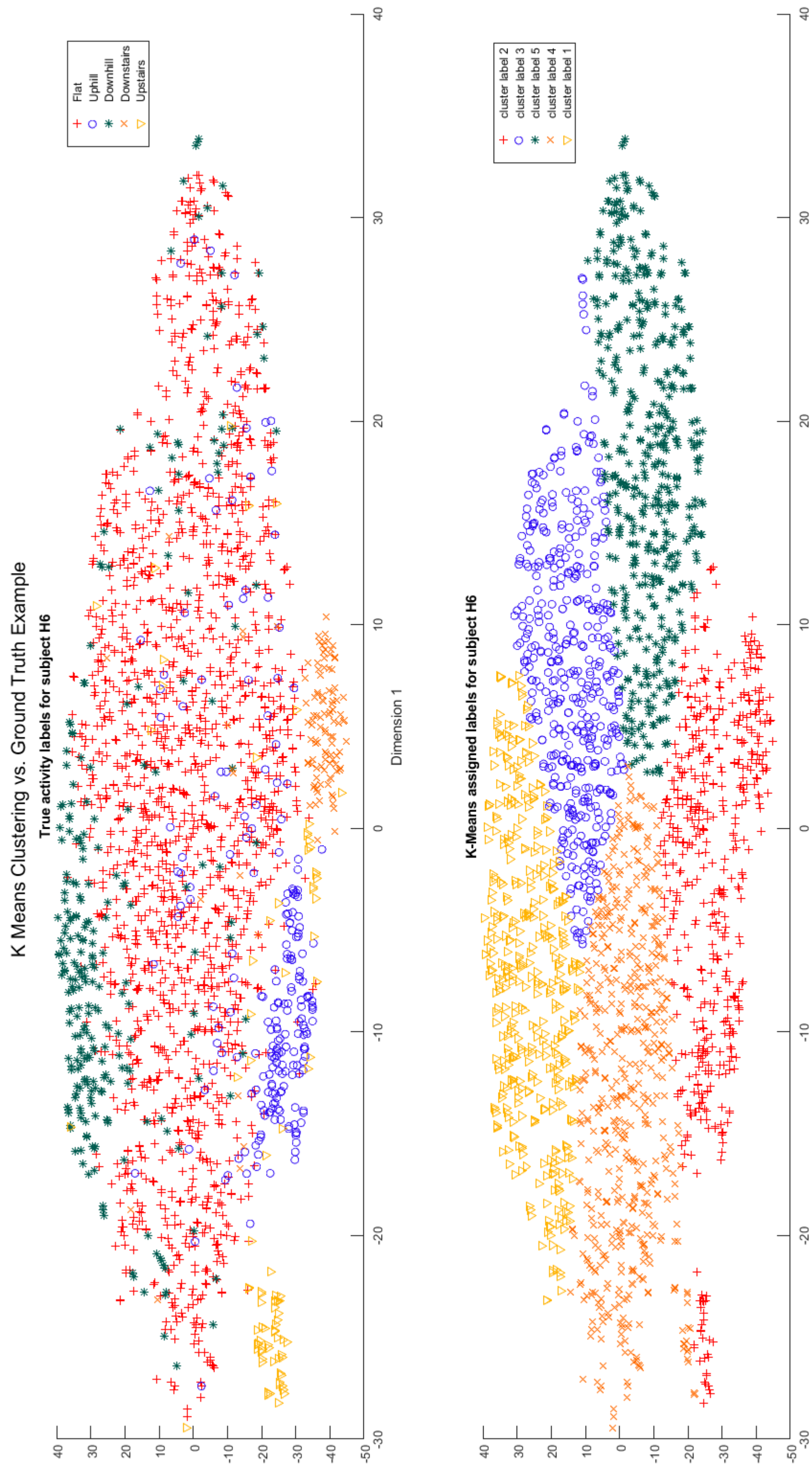


Figure 7.34 tSNE-created cluster plot of amputated individuals for terrain resolution Level 2

Due to the general inseparability of data at Level 1 of terrain resolution, a short experiment involving the application of clustering algorithms was carried out to determine the feasibility of clustering at this level of resolution. K-Means, Hierarchical and GMMs were employed on each subject. To ensure that all clustering models were optimized, a grid-search algorithm combined discrete values of hyperparameters and were tested on each subject for 5 different initiations of the t-SNE algorithm. The overall best clustering parameter combinations for all subject and all t-SNE initiations was then tested on another five tSNE initiations for each subject. For validation, the mean of the Normalized Mutual Information (NMI) across the five initiations for each subject was chosen. NMI is one of a handful of external validation measurements used in unsupervised learning, where a ground truth annotation of the clusters can be provided. Mutual Information is an entropy measurement that measures how closely the designated cluster labels resemble the ground truth, and normalizing the value ensures that the range of NMI falls between 0 (no resemblance) to 1 (perfect resemblance). NMI is preferred over a similar comparison metric, the Adjusted Rand Index, when the distribution of cluster sizes is unequal (Romano et al., 2015). The resultant NMIs are displayed in Table 7.1. These results demonstrate that the 3 most well-known types of clustering algorithms are unable to form meaningful clusters that relate to physical activity. This is further evidenced by Figs. 7.35-7.37, where the clusters formed by these algorithms do not resemble the ground truth.

Table 7.1 NMI of cluster label assignment vs. ground truth for various clustering labels

Classifier	K-Means	GMM	Hierarchical
H1	0.155±0.006	0.143±0.025	0.122±0.023
H2	0.246±0.020	0.224±0.032	0.224±0.016
H3	0.138±0.007	0.133±0.014	0.128±0.011
H4	0.173±0.012	0.188±0.009	0.148±0.018
H5	0.133±0.017	0.155±0.011	0.150±0.018
H6	0.203±0.009	0.243±0.013	0.159±0.011
H7	0.114±0.013	0.106±0.014	0.132±0.023
H8	0.069±0.003	0.071±0.003	0.075±0.004
A1	0.165±0.003	0.186±0.012	0.130±0.010
A2	0.080±0.003	0.089±0.006	0.075±0.003
A3	0.145±0.013	0.190±0.024	0.123±0.012
A4	0.156±0.013	0.182±0.012	0.144±0.016
Mean	**0.147±0.049**	**0.157±0.053**	**0.134±0.039**
Best Parameters	**Distance Metric: Squared Euclidean**	**Probability Tolerance: 1e-8, Regularization Value: 1**	**Distance Metric: Cosine, Linkage: Ward**



K Means Clustering vs. Ground Truth Example
True activity labels for subject H6
Flat
Uphill
Downhill
Downstairs
Upstairs
Dimension 1
Dimension 2
K-Means assigned labels for subject H6
cluster label 2
cluster label 3
cluster label 5
cluster label 4
cluster label 1
Dimension 1
Dimension 2

Figure 7.35 Top: Cluster model of participant H6 after tSNE dimensionality reduction, labelled by its ground truth activity labels. || Bottom: Cluster model of participant H6 after tSNE dimensionality reduction, labelled by K-Means clustering algorithm assigned labels.

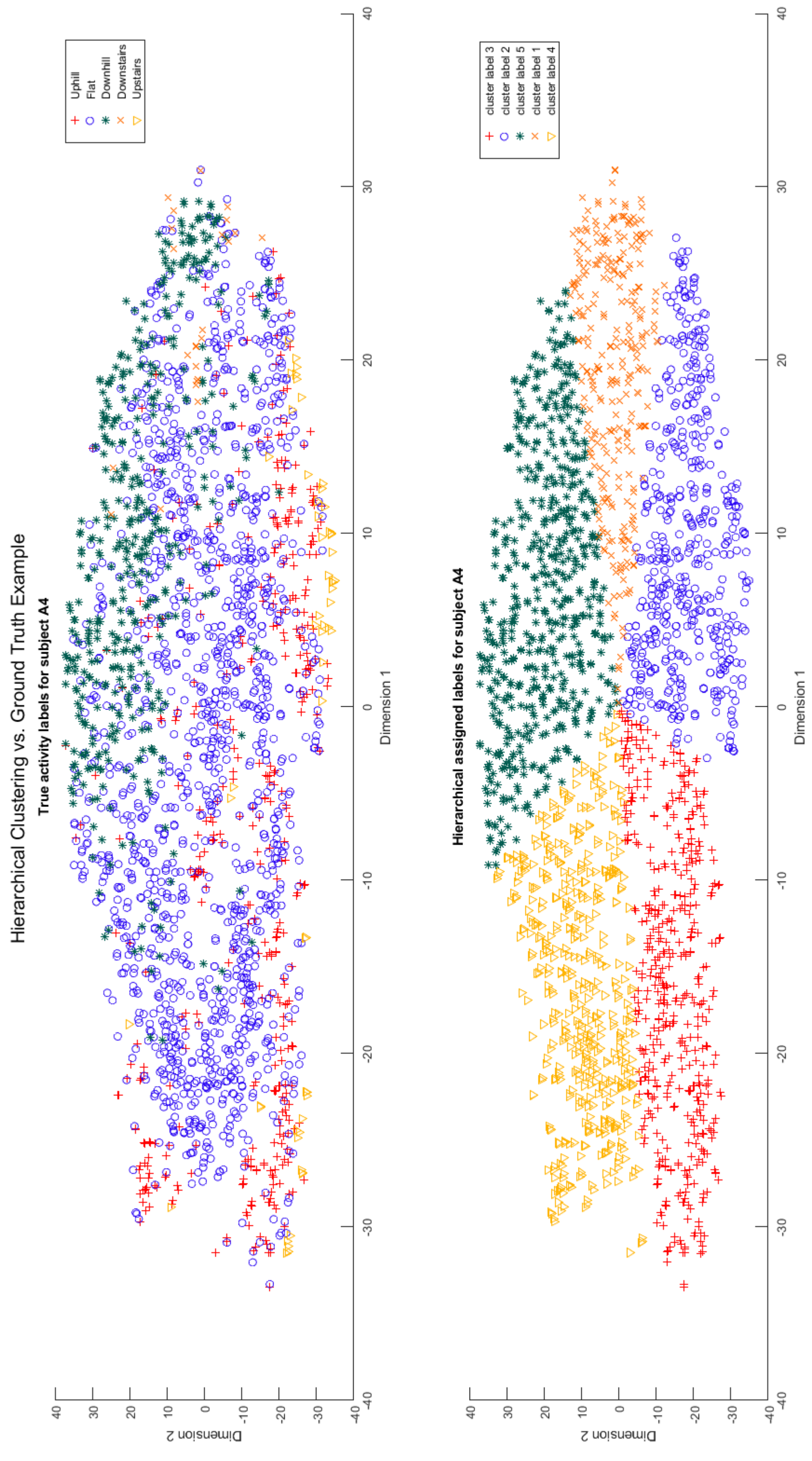

Hierarchical Clustering vs. Ground Truth Example
True activity labels for subject A4
Uphill
Flat
Downhill
Downstairs
Upstairs
Dimension 1
Dimension 2
Hierarchical assigned labels for subject A4
cluster label 3
cluster label 2
cluster label 5
cluster label 1
cluster label 4

Figure 7.36 Top: Cluster model of participant A4 after tSNE dimensionality reduction, labelled by its ground truth activity labels. || Bottom: Cluster model of participant A4 after tSNE dimensionality reduction, labelled by Hierarchical clustering algorithm assigned labels

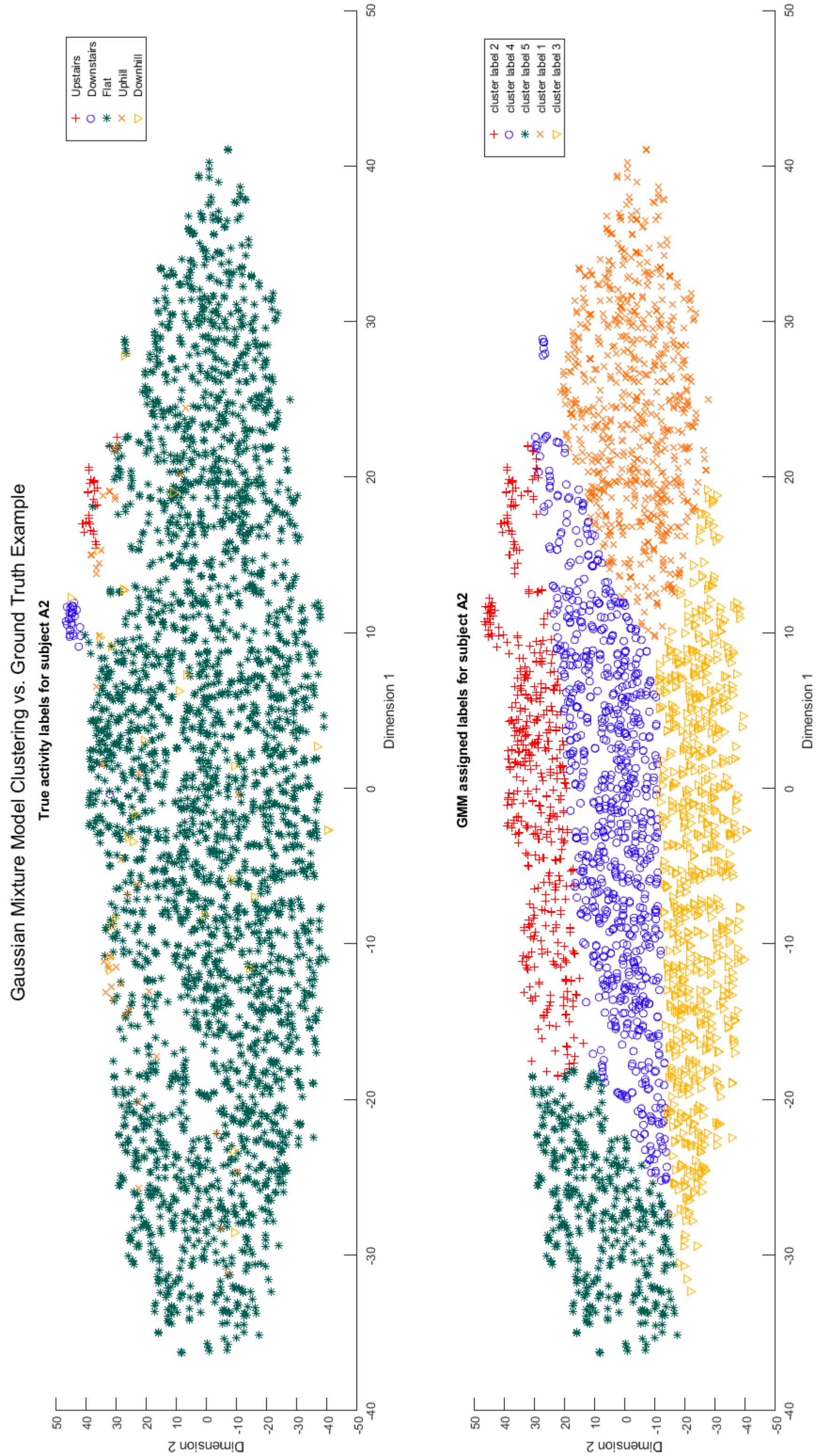

Gaussian Mixture Model Clustering vs. Ground Truth Example
True activity labels for subject A2
Upstairs
Downstairs
Flat
Uphill
Downhill
Dimension 1
Dimension 2
GMM assigned labels for subject A2
cluster label 2
cluster label 4
cluster label 5
cluster label 1
cluster label 3
Dimension 1
Dimension 2

Figure 7.37 Top: Cluster model of participant A2 after tSNE dimensionality reduction, labelled by its ground truth activity labels. || Bottom: Cluster model of participant A2 after tSNE dimensionality reduction, labelled by GMM clustering algorithm assigned labels.

In reality, at Level 1 of label resolution, the amount of overlap in walking, downhill and uphill labels simply cannot be separated in a practical manner for the majority of the subjects, no matter how sophisticated the clustering algorithm is. This is most apparently evident by looking at the true activity labels for subject H3 in Fig.7.38. As a result, the investigation pivoted towards studying for clustering activity levels at a much simpler level: distinguishing walking (flat, uphill, and downhill) from stairs (up and down). As illustrated in Figs. 7.35-7.37, most of the included participants had appreciable separable clusters for stairs, and thus had a chance of generating viable clustering algorithms.

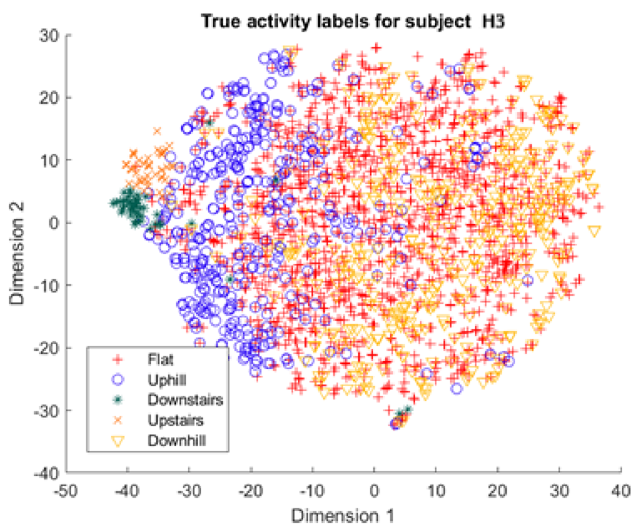


Figure 7.38 tSNE cluster model of participant H3, labelled by ground truth activities.

7.2.6 Exercise #4: Tuning Dimensionality Reduction Method Parameters

At resolution level 0, model parameters for the tSNE algorithm were manually adjusted. tSNE hyperparameters include the number of dimensions (2 or 3), whether PCA is applied to the dataset prior to the tSNE transformation, the distance metric for calculating similarities, the perplexity (number of effective nearest neighbours), and exaggeration (the size of natural clusters). The tSNE parameters were tuned one parameter at a time, using all 12 subject cluster models and visually comparing them. In general, the choice of parameter had minimal visual improvement on the cluster models over the default configurations, and the models tended to fail to form meaningful clusters when parameters were set to their extreme limits. A configuration of numDimensions = 2, perplexity = 30, exaggeration = 4 and distance = “Standardized Euclidean” was found to be an appropriate solution for all clustering models. Applying PCA prior to tSNE also resulted in poorer cluster models for various levels of principal component preservation, and so it was decided against using PCA.

After completion of exercise #4, it was established that any unsupervised clustering models would, at best, be able to distinguish between walking and ascending/descending stairs, and so the objective A was complete. The second part of the methodology focuses on objective B, which is to form a clustering algorithm that can capture these activities in a cluster and correctly recognize them.

7.3 Part II: Constructing a Clustering Algorithm

With the absorption of uphill, downhill, and level walking movement into a single label at Level 0 resolution, there was a large class imbalance, with Walking data points typically encompassing between 90 to 99% of the dataset for an individual. As the results from exercise #4 demonstrated, parametric clustering algorithms where the number of clusters is pre-determined were an ill-fit for cluster recognition, as they typically tended towards forming clusters with equal sizes; these findings are reflected in the literature: Franti and Sieranoja (2019) demonstrated that, regardless of the cluster initialization method, their K-means algorithm performed poorly on unbalanced datasets. The solution was to employ nonparametric clustering algorithms (algorithms that do not predetermine the number of clusters) that were better suited towards recognizing clusters of different sizes (Finch, 2019). There are a number of non-parametric clustering methods, including OPTICS, Dirichlet Gaussian Mixture Models and Hierarchical Clustering (which can be both parametric and non-parametric) (Ankerst et al., 1999; Görür and Edward Rasmussen, 2010; Mohamad, Bouchachia, and Sayed-Mouchaweh, 2015). Upon reflection of the methodology of Chapter 6, it was determined that more meaningful results would be obtained by focusing on optimizing a single non-parametric clustering technique, hence it was decided to focus using clustering only with DBSCAN. DBSCAN's operation was previously explained in Chapter 4.2.6.2.3.

7.3.1 Exercise #5: Clustering with DBSCAN

The DBSCAN algorithm underwent the same validation process as the other three clustering algorithms in exercise #3: five tSNE models for each person were created, and the NMI was calculated for discrete values of DBSCAN parameters (epsilon and minimum number of points for cluster membership). The label resolution is set to level 0, such that it only includes three activities: “Walking”, “Upstairs” and “Downstairs”. Upon completion, the highest scoring DBSCAN parameters were applied to another 5 tSNE models, and the average NMI for each subject was subsequently calculated. The results are demonstrated in Table 7.2. An example of the resulting DBSCAN clusters is shown in Fig. 7.39.

Comparing the results from Table 7.1, DBSCAN had significantly better NMI compared to K-Means, Hierarchical and GMM, though in part this was due to the reduced level of label resolution. It was evident the DBSCAN algorithm was good at acquiring stair data when

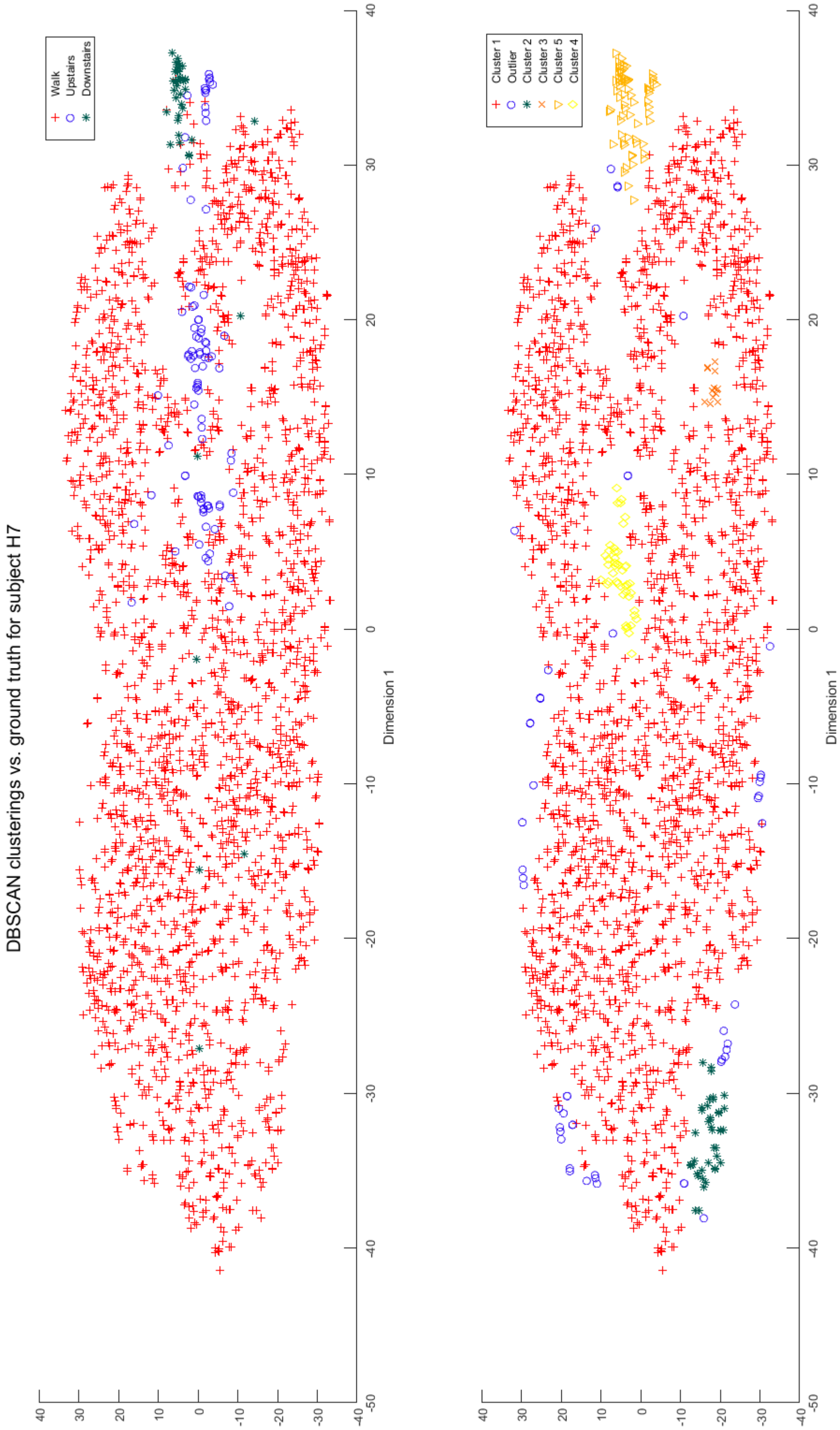

DBSCAN clusterings vs. ground truth for subject H7
Walk
Upstairs
Downstairs
Cluster 1
Outlier
Cluster 2
Cluster 3
Cluster 5
Cluster 4
Dimension 1
Dimension 2

Figure 7.39 Top: Cluster model of participant H7 after tSNE dimensionality reduction, labelled by its ground truth activity labels. || Bottom: Cluster model of participant H7 after tSNE dimensionality reduction, labelled by DBSCAN clustering algorithm assigned labels.

the tSNE model successfully clusters them; in Fig. 7.39, DBSCAN is able to capture the downstairs cluster for subject H7. However because the upstairs data has not successfully clustered, the DBSCAN algorithm has no reasonable ability to recognize the upstair sample points as a separate cluster. The immediate disadvantage of the DBSCAN approach is that because the number of clusters cannot be predetermined, the DBSCAN algorithm tends to form additional "false" clusters that simply contain more walking data. To combat this problem, a specialized algorithm was developed.

7.3.2 Development of a "Walk" And "Stair" cluster recognition algorithm

By reducing the classification problem to a binomial outcome (a cluster is "walk", or a cluster is "stairs"), an effective cluster recognition algorithm was developed without the need of using supervised training data. A psuedo-code summary of the algorithm is presented in Appendix D. Based on observations of the true labels in the cluster models, the following hypothesis was made:

I. The largest cluster identified by DBSCAN is always a walking cluster

Table 7.2 NMI of subject models after DBSCAN cluster assignment vs. ground truth

Subject	H1	H2	H3	H4	H5	H6	H7	H8	A1	A2	A3	A4
Mean NMI	0.239	0.279	0.010	0.266	0.419	0.377	0.226	0.149	0.538	0.595	0.168	0.023
Std.	±0.030	±0.176	±0.002	±0.022	±0.091	±0.125	±0.026	±0.046	±0.07	±0.057	±0.038	±0.044
Best Parameters: Epsilon = 3, Minpts = 15												

Hypothesis I can be validated without the need for statistical analysis. It can be observed in all DBSCAN-related figures in this chapter that the largest DBSCAN-identified cluster contains an overwhelming majority of “walk” data points when compared to the ground truth label plots. Realistically, there is no practical free-living scenario in which a subject will have more than 20% of their total data be based on stairs. Thus the only scenario in which a stair cluster could be the largest DBSCAN cluster is if the volume of input data is too sparse, resulting in many incidental DBSCAN clusters being created. The outlying points detected by the DBSCAN algorithm can also be eliminated from consideration, either as a walking cluster or a stair cluster. If after applying the DBSCAN algorithm only one cluster has been identified (excluding errors), the algorithm throws an error.

Proceeding the identification of the primary walking cluster, there will remain a number of other clusters detected by DBSCAN, and the objective is now to determine which of these DBSCAN-detected clusters is the “walking” cluster. If one other DBSCAN cluster has been identified (excluding the primary walk cluster and outlier points), this cluster is automatically assigned to be the stair cluster by simple process of elimination. If there are two or more clusters however, then the algorithm must identify the DBSCAN-cluster with the highest probability of being a stair cluster. The assumption of the algorithm is that only one of these clusters is the stair cluster. All other DBSCAN-detected clusters are assumed to be another “walk” cluster that was not recognized as part of the primary walking cluster.

It can be deduced through simple probability theory that the DBSCAN-detected cluster C_i that is most likely to be the stair cluster C_{stairs} is the one that is least likely to be a walking cluster C_{walk}:

$$P\left(C_i = C_{stairs}\right) = 1 - P(C_i = \ C_{walk}) \tag{7.1}$$

Through further observation of the stair clusters in the tSNE models, 2 more hypotheses were formed:

II. Stair clusters tend to be the cluster located the furtherest (in terms of Euclidean distance) from the walking cluster

III. Stair clusters tend to have more compact clusters than walking clusters

For hypothesis II, the probability of a DBSCAN-detected cluster being the stair cluster can be calculated as the product of distance of the cluster's centroid from the main walking cluster. To calculate this factor, a GMM with 1 component is fitted to the main walking cluster, thereby forming a gaussian distribution around its centroid. Fig. 7.40 gives an example illustration of the GMM with a probability density function emanating from the centroid of the walking cluster.

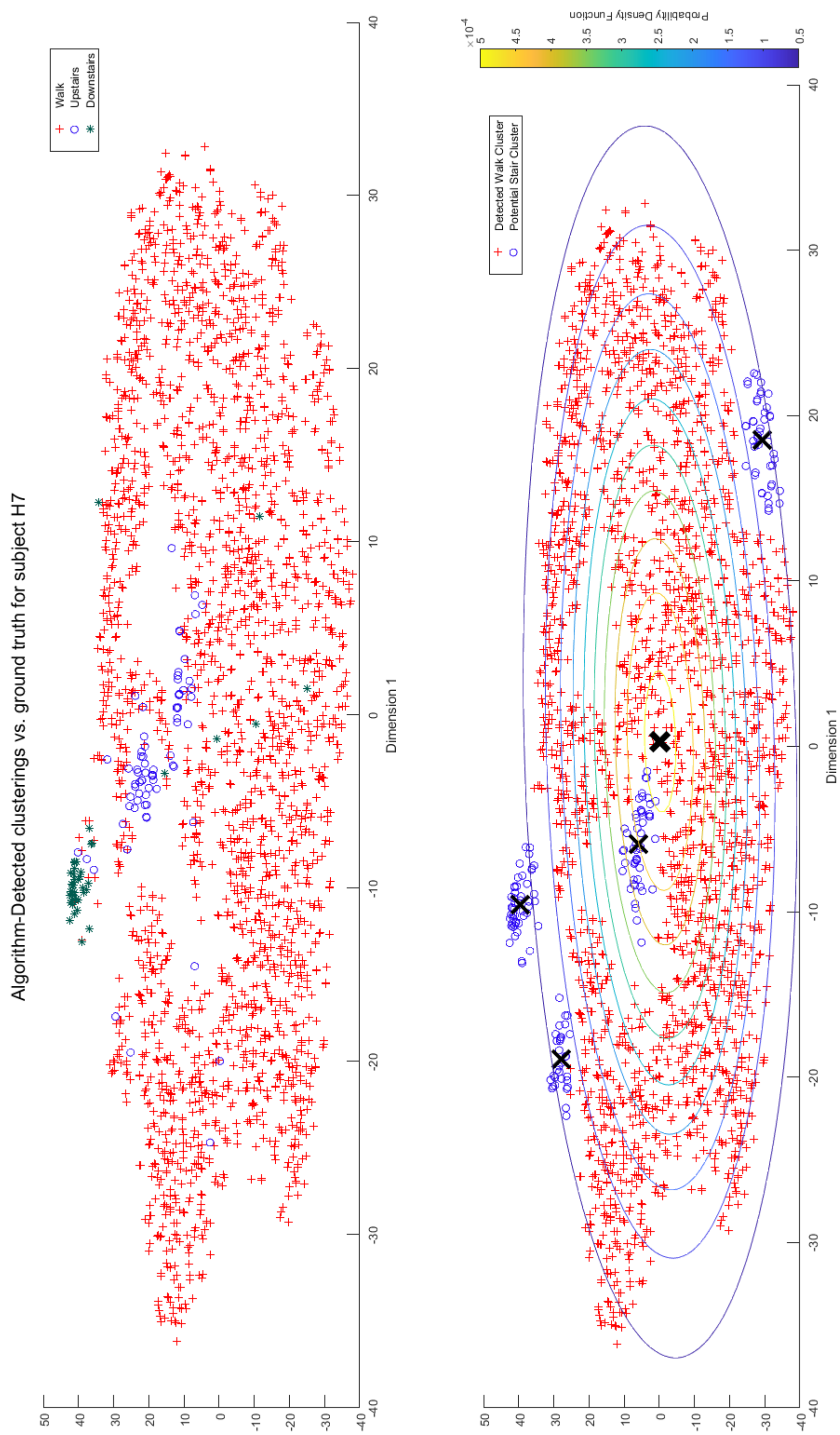

Algorithm-Detected clusterings vs. ground truth for subject H7
Walk
Upstairs
Downstairs
Dimension 1
Dimension 2
Detected Walk Cluster
Potential Stair Cluster
Probability Density Function

Figure 7.40 Top: tSNE cluster model of participant H7 labelled with ground truth activities || Bottom: Initial application of the proposed algorithm, with recognition of a primary 'walk' cluster and several potential 'stair' clusters.

The resulting negative log-likelihood of the GMM given the centroid data of each DBSCAN-detected cluster (excluding the already recognized main walking cluster) can then be calculated using posterior probabilities. The negative log-likelihood, in the context of machine learning, is a cost function of how well a data point or set of data points fit to a machine learning model. The lower the negative log-likelihood, the greater the fit. Thus, clusters with centroids that are located far from the main walking cluster will have high negative log-likelihoods, this will make the data point a poor fit for the walking cluster thereby being a better fit for a stair cluster.

For hypothesis III, the compactness of a DBSCAN-identified cluster is simply determined as the median of the pairwise distances between all observations. The median was chosen as the summative property to mitigate the effect of any outlier observations in the cluster.

Using these two hypotheses, an equation for stair cluster probability was composed (Eq. 7.2). It calculates the probability of a DBSCAN-detected cluster, C_i being the stair cluster C_{stairs} as follows:

$$P\left(C_i = C_{stairs}\right) = 1 \ - \ \frac{1}{\mathcal{L}\left(K_{c_i} \in \ C_{walk}\right) \ \times \ \varphi_{j(C_i)}} \tag{7.2}$$

$\mathcal{L}\left(K_{c_i} \in \ C_{walk}\right)$ represents the negative log-likelihood that a centroid of cluster C_i belongs to the walking cluster C_{walk}. $\varphi_{j(C_i)}$ represents a ranking factor of the clusters compactness. For a single model, the cluster with the least amount of compactness is given a value of one. The cluster with the next highest level of compactness is given a value of one plus a scalable reward factor, repeating for all remaining clusters up to the highest compact cluster. Once the two main factors have been calculated, the probability of each cluster C_i for all i clusters are calculated, and the cluster with the highest probability is assigned as the stair cluster. All other remaining clusters are reassigned as a walking cluster.

The differences of negative log-likelihood and compactness between stair clusters and "extraneous walking" clusters (i.e DBSCAN-detected Walk clusters that are smaller than the largest walking cluster) are not vast enough such that some statistical analysis was required to validate the hypotheses. Given the small number of participants, the number of clusters available for analysis was fairly small. Fortunately, the randomness of the tSNE model building process can be taken advantage of to help combat the small dataset sizes: by setting Matlab's random number generator seeding to a fixed value, the resulting tSNE model will be consistently the same when the script is ran multiple times. The RNG seed was varied from 1 to 5, and the resulting cluster compactness and negative log-likelihoods from each tSNE model was acquired. The statistical analyses of the observations II and III are discussed in the results.

7.3.3 Algorithm Validation

For each subject, the clustering algorithm is evaluated on three factors:

1. The DBSCAN success rate
2. Stair cluster purity
3. Algorithm-corrected NMI

The success rate is a basic measurement of whether the DBSCAN algorithm was successful in identifying at least 2 but no more than 10 clusters. Cluster purity is acquired through reading the labels of the data points of the cluster that has been assigned as the stair cluster, and subsequently calculating the proportion of the number of labels that are either “upstairs” or “downstairs”. Once clusters that do not have the highest stair cluster probability are reassigned to have the same cluster identity as the main walking cluster, the resulting algorithm-corrected tSNE model will only have 2 clusters. This model can then be compared to the ground truth labels and assessed via NMI. The process is ran for 5 unique tSNE model iterations for each subject.

7.4 Results

7.4.1 Validation of Hypotheses

From 5 iterations of tSNE models for each subject, a total of 55 stair clusters and 45 extraneous walking clusters were acquired. From each cluster, the compactness and negative log likelihood of fitting to the main walking cluster of the respective subject was calculated.

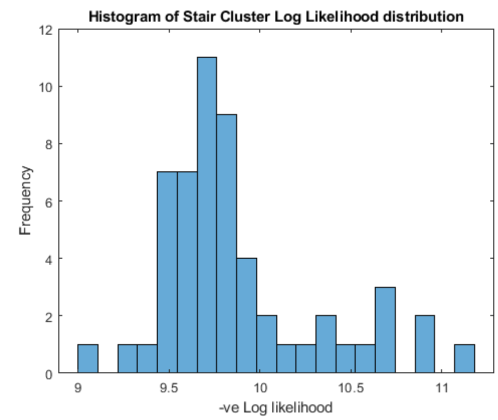
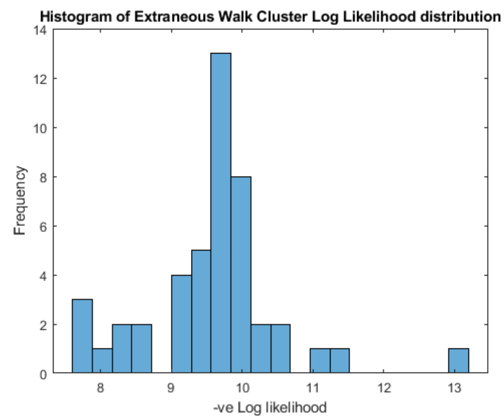
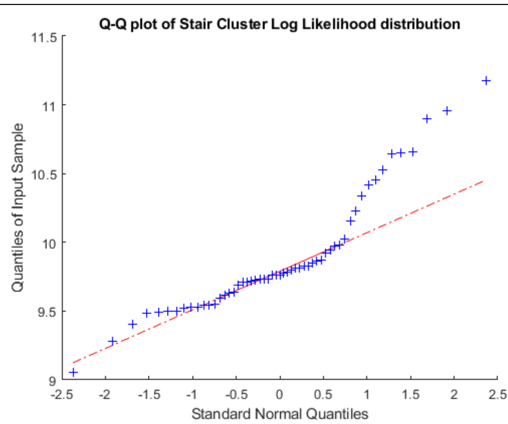
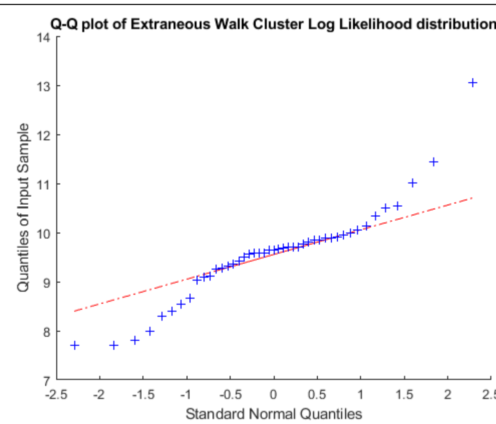
The first stage in statistical analysis was a check for normal distribution, which was achieved through a qualitative and quantitative combination of Q-Q plots, histograms, and the Shapiro-Wilk test (Shapiro and Wilk, 1965). The null hypothesis of the Shapiro-Wilk test states that the distribution from a sample X follows a normal distribution, and that composite normality for the sample can be assumed. Thus, a p value significant enough to reject the null hypothesis ($\alpha = 0.05$) indicates that the sample has another type of distribution.

Tables 7.3 and 7.4 demonstrate that both the compactness and the log-likelihood of the stair and walking clusters do not follow normal distribution. The histograms cannot be approximated with a bell-shaped curve, and the Q-Q plots do not fit a linear gradient.

As sample distributions were abnormal, Student's T-test was considered inappropriate for testing the equivalence of each sample's means. In place, the Wilcoxon rank-sum test was applied. The Wilcoxon rank-sum's null hypothesis differs from the T-test in that it assumes that the probability of one variable, X, being greater than another variable, Y, is equal to the probability of Y being greater than X (Wilcoxon, 1945). For the purposes of this investigation, this is a suitable metric to test for equality between the stair and extraneous walk clusters (Fay and Proschan, 2010). The significance level is set to 95% ($\alpha = 0.05$).

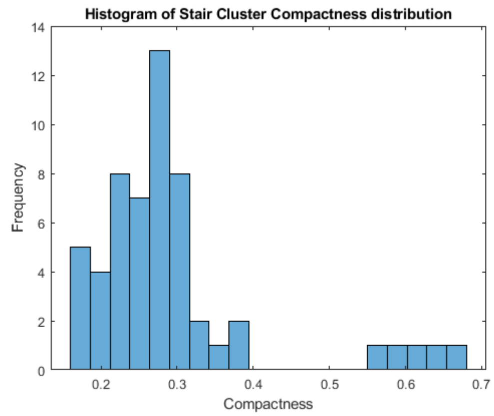
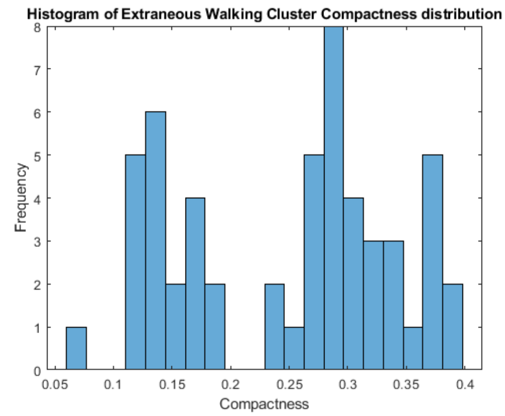
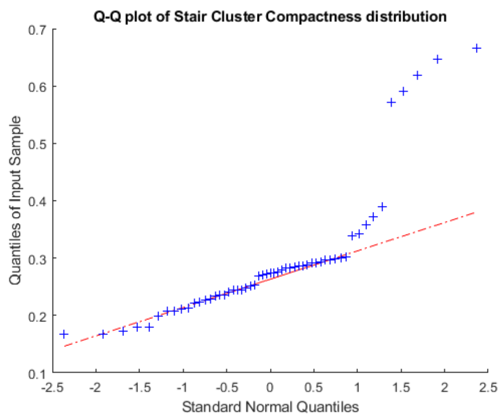
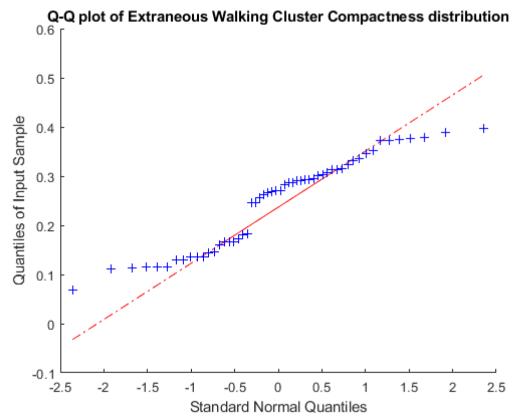
From Table 7.5, the null hypothesis for the negative log-likelihood could be rejected;

Table 7.3 Testing for normal distribution for cluster negative log-likelihood

<table>
<tr><th>Normality Test Method</th><th>Stair Clusters</th><th>Extraneous Walk Clusters</th></tr>
<tr><td>Histogram</td><td>
</td><td>
</td></tr>
<tr><td>Q-Q plot</td><td>
</td><td>
</td></tr>
<tr><td>Shapiro-Wilk p value</td><td>1.70e-4†</td><td>9.46e-4†</td></tr>
</table>

† = statistically significant

Table 7.4 Testing for normal distribution for cluster compactness

<table>
<tr><td>Normality Test Method</td><td>Stair Clusters</td><td>Extraneous Walk Clusters</td></tr>
<tr><td>Histogram</td><td>
</td><td>
</td></tr>
<tr><td>Q-Q plot</td><td>
</td><td>
</td></tr>
<tr><td>Shapiro-Wilk p value</td><td>1.22e-7†</td><td>0.003†</td></tr>
</table>

† = statistically significant

Table 7.5 Testing for statistical significance in negative log-likelihood and compactness for Stair and Extraneous Walk clusters.

Property	Negative Log-likelihood	Compactness
Rank-sum p-value	**0.023**†	0.387

† = statistically significant

Table 7.6 Success rate of DBSCAN clustering

Subject	H2	H3	H4	H5	H6	H7	H8	A1	A2	A3	A4
Mean Success Rate(%)	80	80	100	100	100	100	100	100	100	100	40

however, the compactness could not. This would imply that only the distance of the cluster centroid from the main walking cluster is the only significant factor in whether the cluster is a stair cluster or extraneous walk cluster. It was not desirable to have Eq. 7.2 be univariate, and considering that the compactness p value was significant when a t-test was applied (p = 0.026), a compromise was made by making cluster compactness a very weak weighting factor: each cluster in successive order of compactness had an additional rank factor ($\varphi_{j(C_i)}$) of 0.05.

7.4.2 Clustering Algorithm Results

The sparsity of data in non-amputated subject 1 ($<$ 1000 samples of data) led to their exclusion from the validation process. Referring back to Chapter 5.2.4, this was due to two of their recordings being discarded due to improper camera wear. Table 7.6 demonstrates the success rate of being able to detect between 2 and 10 clusters from the resulting tSNE model after outliers have been removed. Subject A4 was the only subject to have considerable trouble acquiring more than a single DBSCAN cluster.

Table 7.7 presents the purity of stair data counts found within the detected stair clusters. The results from each iteration have been included to highlight the inconsistency between tSNE models. When the stair cluster was correctly recognized however, the purity was relatively consistent. There was clear failure to recognize stair clusters in H3 and A4, which reflects on the poor quality of their original tSNE models. Appendix E contains exemplary comparisons of the ground truth activity labels for each subject and comparisons with the algorithm-detected cluster labels.

Table 7.7 Table of cluster purity of stair labels in identified stair clusters, expressed as %

Subject	H2	H3	H4	H5	H6	H7	H8	A1	A2	A3	A4
Model #1	NaN	0	85.11	81.67	100.00	72.22	9.59	69.33	93.33	78.57	NaN
Model #2	72.00	0	85.11	85.96	100.00	78.13	10.34	69.86	93.33	79.82	NaN
Model #3	0	NaN	85.11	84.21	0.00	81.25	9.26	70.83	84.85	79.13	NaN
Model #4	0	0	85.11	0.00	100.00	58.06	94.12	71.23	79.49	80.18	47.06
Model #5	80.18	0	85.11	78.13	96.30	72.22	8.96	69.33	83.10	38.54	0
Mean	**38.05**	**0.00**	**85.11**	**65.99**	**79.26**	**72.38**	**26.45**	**70.12**	**86.82**	**71.25**	**9.41**
Std.	±38.15	±0.00	±0.00	±33.10	±39.66	±7.96	±33.84	±0.78	±5.59	±16.36	±18.82

NaN = "Not a Number", indicating the algorithm failed to recognize more than 1 cluster

The NMI collected after the algorithm has recognized the stair cluster was compared with the initial NMI obtained (identical to those in exercise #5). While the NMI is still quite low in most subjects, the algorithm corrected NMI was found to be greater or equal to the original NMI in all but 2 of the subjects.

Table 7.8 NMI before and after application of proposed algorithm.

Subject	H2	H3	H4	H5	H6	H7	H8	A1	A2	A3	A4
Initial NMI	0.279	**0.010**	0.266	0.419	0.377	**0.226**	0.149	**0.538**	0.595	**0.168**	0.023
Std.	±0.176	±0.002	±0.022	±0.091	±0.125	±0.026	±0.046	±0.070	±0.057	±0.038	±0.044
Final NMI	**0.374**	**0.010**	**0.367**	**0.433**	**0.377**	0.216	**0.164**	**0.538**	**0.608**	0.155	**0.027**
Std.	±0.133	±0.001	±0.048	±0.086	±0.125	±0.017	±0.078	±0.046	±0.064	±0.052	±0.052

To reinforce the positivity of the findings from the algorithm, an additional right-tailed pairwise t-test was carried out to determine, on average across all subjects (excluding H1), whether the algorithm corrected NMI was greater than the initial NMI after only performing DBSCAN, with the null hypothesis stating that there was no significant difference. The null hypothesis was rejected ($p = 0.0031$) indicating that the proposed algorithm made a systematic improvement over only using DBSCAN.

7.5 Discussion

7.5.1 Clinical Significance

Overall, the experience of attempting to provide a clinically beneficial analysis of free-living activity data of non-amputated and lower limb individuals with an unsupervised approach proved to be challenging. Recalling back to Fig.7.13, it was found that when the population model contained multiple participants, the non-linear dimensionality reduction techniques formed clusters based on the individual rather than the activity. This phenomenon was especially pronounced in the ILLA population model (Fig. 7.13 B & D). This is in contrast to other unsupervised HAR studies with non-amputated populations, such as with Kimel-Naor, Gottlieb, and Plotnik (2017) who were able to cluster activities with multiple participants in the Opportunity dataset, implicating that the uncontrolled environment for recording data led to a greater diversification of activities that were unique to each individual. Within the ILLA sample group, there was one bilateral amputee and three unilateral transtibial amputees. Accounting just for the three transtibial amputees, there was some diversity in regard to their age and prosthetic experience, both of which can have an impact on gait patterns (Highsmith et al., 2010; Vanicek et al., 2009). Given the very small sample size, it is difficult to comment on how the modelling would change if more participants could have been recruited. By extension, “training” an unsupervised model on non-amputated participants and testing on an ILLA population would also be unfeasible given differences in gait behaviour (Lemaire and Fisher, 1994; Su et al., 2008). Regardless, the findings of this investigation imply that an unsupervised approach is only viable by analysing data from the individual, and not from a general population.

Further to this, when the true activity labels were investigated, it was found that discernible clustering only existed between stair activity and walking activity, with walking activity encompassing level walking, uphill and downhill movement. While a few subjects (subjects H2 and H6 for instance) showed an appreciable linear gradient-like distribution of flat, uphill, and downhill movement in the main walking cluster, the majority of subjects did not display this behaviour. In Fig. 7.41 for subject H2, it can be seen that, ignoring outliers, the downhill labels are distally the furthest away from upstairs and downstairs activities, flat labels are slightly closer, and uphill labels are the closest. These differences may arise as a result of different natural walking speeds in level walking, hills, and stairs. Unfortunately, it was not possible to validate the walking speed in this study, but gait kinematics literature appears to support this notion. From Sun et al. (1996) it was observed that in young adults, walking speed generally tends to increase as the angle of slope decreases. Likewise, it decreases, though to a lesser extent, when traversing uphill. Fujiyama and Tyler (2004) demonstrates that walking speeds on flat surfaces are significantly faster than walking speeds when ascending or descending stairs, even when stairs were being descended quickly, and was true for both a young adult and elderly population. In ILLA populations, significant differences in preferred walking speed were observed by Rodrigues, Andrade, and Vieira (2019) in transtibial and transfemoral amputees for level walking, uphill and downhill movement, though interestingly no significant differences in speed were observed in the non-amputated control group. Finally, Wolf et al. (2012) illustrated significant differences in walking speed between ascending ramps and ascending stairs, and descending ramps and descending stairs in ILLAs. These sources indicate that, on a fundamental level, there will be some change in walking speed between the five main types of ambulation. Therefore, this investigation hypothesizes that successful clustering of stair and hill movement is largely dependable on appreciable changes in walking speed. As the walking speed changes, features relating to the walking speed such as energy of acceleration, the integral features sourced from Rosati, Balestra, and Knaflitz (2018) and frequency related features will have significantly different values, which the tSNE algorithm is able to recognize and cluster together in low dimensions. In stair movement, this change in walking speed should be much more appreciable as the subject has to decrease their walking speed to traverse the stairs in a safe manner. Whereas in hill movement, if the angle of the slope is not steep, the subject does not need to make significant alterations to their walking speed, thereby resulting in greater interspersion of the hill and level walking data in the tSNE models. Naturally, this theory will require further testing with validation of the walking speed. For cluster models that have gradient distribution of walking activities, there could be potential for accurate clustering via innovative novel clustering techniques like Convex Clustering (Chi and Lange, 2015). Moreover, the addition of the magnetometer features in the study appears to help increase the quality of the models, which is investigated in Chapter 8.

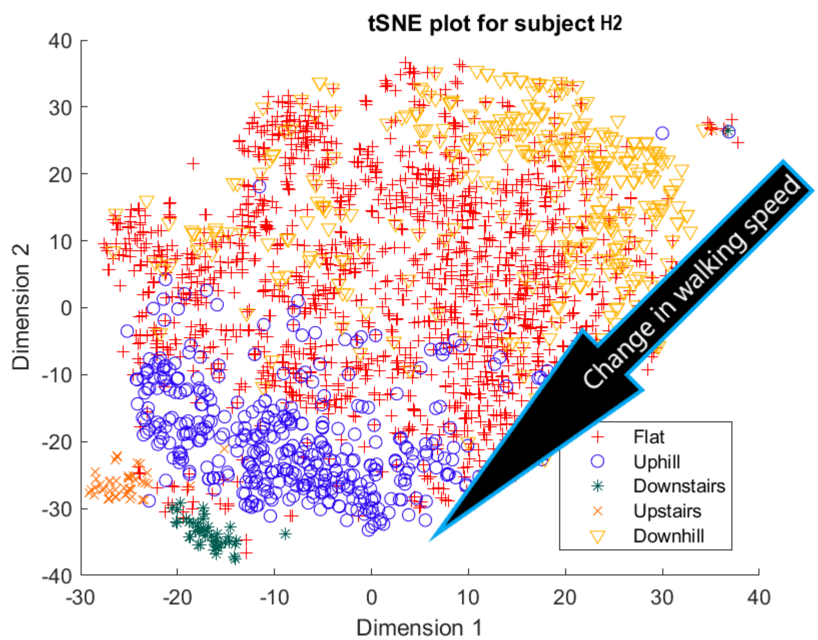


Figure 7.41 tSNE cluster model of participant H2, with an arrow indicating general change (decrease) of walking speed in the 2 dimensions

Despite being only able to distinguish two kinds of activities ("walking" and "stairs"), the implications of the investigation could be beneficial when applied in the early stages of rehabilitation of an ILLA. A patient, who may have recently acquired a prosthetic fitting for the first time, or changing to a new prosthetic component, may have trepidations about traversing stairs. Recalling back to Chapter 3.2.2.4, even the interviewed ILLAs, who had at least several years of experience with a prosthetic fitting, were still hesitant about going downstairs. From that perspective, providing a system in which stair traversal activity can be recognized without the need of training or annotating data could be useful for a healthcare professional who wanted to evaluate the progress of their clients. A caveat is that, as per the limitations of the DBSCAN algorithm, there would need to be a reasonable amount of stair data collected for the DBSCAN algorithm not to accidentally misattribute the data as outlier points.

Regardless of the findings from the investigation, further experimentation on an ILLA population, particularly those in the early stages of rehabilitation, is warranted in order to fully comment on the capability of the unsupervised learning approach. If an early-stage ILLA is having difficulties in traversing sloped environments, this may also result in separable walking clusters via marked changes in walking speed and gait patterns. Intrinsic cluster evaulation methods such as the Silhouette method were not considered for integration into the methodology (Rousseeuw, 1987), due to having the ground truth available and also due to requiring specific demands of the cluster behaviour (i.e they must cluster based on activity). However, future research should incorporate the Sihlouette score as an evaluation of cluster density, and the score may also provide an additional criteria point for identifying stair clusters.

7.5.2 Comparisons with Similar Work

Studies that are most comparative with this investigation are those that have attempted to differentiate stair and walking activities using only wearable sensors through an unsupervised method, generally for a non-amputated population. Unfortunately, there appear to be only a handful of studies that have attempted this. Huynh (2008) was able to distinguish walking and stair movements (upstairs and downstairs separately) with high precision and recall values for each activity. Their approach to unsupervised learning differs in that instead of analysing a low dimensional clustering space, they utilize the eigenspaces generated by feature vectors for unsupervised classification. Their sensory set-up however required a cumbersome and extensive sensory network, with multiple sensors needing to be placed on different locations around the body in order to acquire good performances. Trabelsi et al. (2013) used an unsupervised variant of the Hidden Markov Model known as Multi-Hidden Markov Model Regression to distinguish walking and ascending stairs with high precision and recall values. Like Huynh (2008), their caveat is that data collection required a multi-sensor array worn across the whole body. Both studies also had significantly smaller numbers of participants than was included in this research, limiting their generalizability to a larger and more diverse population. These studies also appear to have more favourable results by using traditionally supervised-based validation metrics (recognition accuracy) over external clustering validation metrics, such as the NMI used in this study. Given the large class imbalance between stair and walking movement, the existence of outliers within in the minority (stair) cluster will proportionately magnify the reduction in mutual information (i.e the shared labels) between calculated and true cluster identities. Finally, both Trabelsi et al. (2013) and Huynh (2008) were carried out in laboratory conditions: this means there was no presence of hill movement (particularly uphill movement) which can potentially obfuscate the cluster separability between strictly flat walking and stair movement. Literature that has attempted to perform clustering while including ramp activities are very rare, likely due to the difficulty of the task. The only identified study to have attempted to do this with wearable sensors appears to be Kafle and Dou (2016), who used EMG sensors and Hierarchical clustering to distinguish level walking, ramps, and stairs. Their highest clustering accuracy achieved was 39.1%, giving a similar performance to this investigation.

7.5.3 Analysis of Confounding Factors relating to Stairs and Cluster Quality

During the walks that participants carried out, it was observed that in general, only a few sets of stairs were traversed per individual. It was then considered a possibility that the number of unique stairs traversed could potentially impact on the clustering model quality: if a person only climbed one set of stairs, all activity data for their stair count would have very similar gait, thereby translating into identical sets of features and allowing the tSNE dimensionality reduction to easily cluster these samples together. The concern was that, if a subject traversed more unique sets of stairs, then the versatility of feature data relating to stair movement increases, thereby diversifying clusters formed in the tSNE process and subsequently adversely impacting on cluster model quality. Another potential confounding factor considered was the quantity of stairs traversed; if cluster model quality is dependent on the number of steps on stairs, then clustering models would naturally be very poor in individuals who have only taken a small number of steps. To investigate these factors, the NMI after cluster adjustment via the proposed algorithm for each individual is ranked in ascending order and independently plotted against the number of unique stairs traversed and total number of step counts. Unique sets of stairs were identified by reviewing the original video footage; note that "unique" refers to the geographic location of stairs: if a person went down one flight of stairs, turned around 180°, and climbed back up those stairs, this would be counted as a single unique set of stairs, regardless of the direction or how many times the set of stairs were traversed. Investigations of similarity were defined through a linear correlation calculation, and the resulting correlation plots are shown in Figs. 7.42-7.43. The R^2 values for the number of stairs and total quantity of steps indicated there was almost zero correlation in regard to the NMI ($R^2 = 0.0005$ and $R^2 = 0.0624$), and therefore the clustering quality is not impacted by these factors. The conclusion is therefore that clustering quality is most likely influenced by appreciable changes in walking speed as a result of ascending or descending stairs. In subjects where cluster quality is poor at level 0, this indicates that the walking speed when traversing stairs is very similar to the walking speed used on flat or hilly terrain.

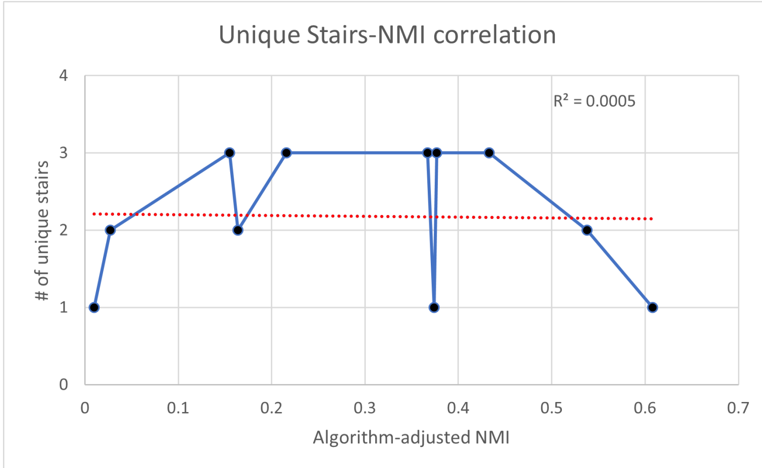


Figure 7.42 Plot of NMI for each participant versus number of unique stairs traversed by the participant

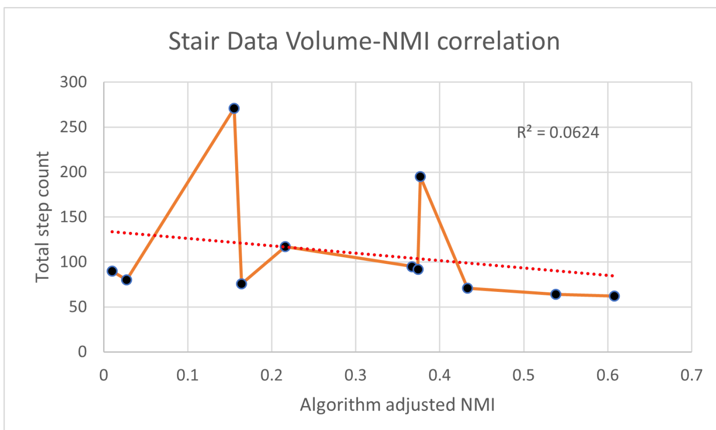


Figure 7.43 Plot of NMI for each participant versus number total number of steps traversed by the participant

7.5.4 Limitations and Future Work

The algorithm presented in this chapter has some obvious weaknesses. It can only detect the presence of a stair cluster, and not whether that movement is upstairs or downstairs. In a clinical context, it could be argued that a healthcare professional may only be concerned whether their client is traversing stairs at all, and the direction of movement may be inconsequential. Nonetheless, future iterations of the algorithm would still benefit from recognizing the direction of stair movement. Another limitation was that, since only one cluster could be assigned as a stair cluster, subjects who achieved two separate clusters for stair movement would have their information misrepresented. However, only 2 out of the 12 subjects exhibited this clustering behaviour consistently (H6 & A2). The majority of subjects would typically only form one stair cluster for one kind of stair movement, while the other stair movement was generally interspersed with the main walking cluster. In a few other subjects (H2, H4 & H5), the upstairs and downstairs clusters were close together enough that DBSCAN would recognize them as a single cluster. Fundamentally, the research should be repeated with a vulnerable ILLA population (i.e those in early stages of rehabilitation) to gain more insight into the clustering behaviour of activities and to test the generalizability of the algorithm. As noted in Chapter 5.9.2, a key factor that limits the generalizability of these findings was that all participants, including ILLAs, used a Step-Over-Step approach to stair ambulation. In a clinical population, ILLAs who have limited prosthetic experience or transfemoral amputation are more likely to use a Step-To ambulation approach to improve their stability in gait amputation (Hobara et al., 2011). Theoretically, this could actually improve the cluster models as the Step-To approach has more differences in gait properties to level walking compared to the Step-Over-Step approach, but would nonetheless require further testing and validation.

Regarding the inconsistency of tSNE modelling, this issue could be fixed through applying a parametric tSNE mapping function. However as explained in the limitations of tSNE models back in Chapter 4.2.5.2.2, this requires training a neural network. As individual models were found to be the only viable way to cluster activities, the tSNE mapping functions could not be effectively utilized given the much smaller dataset sizes.

Developing an unsupervised model within the Matlab environment also highlighted the weakness of Matlab's unsupervised learning capabilities. Official support for clustering algorithms only included the most popular models (i.e K-Means, Hierarchical, Gaussian Mixture Models, Spectral Clustering, and a few others). DBSCAN was the only officially supported non-parametric spatial clustering function. While there were some community-written functions available for comparative non-parametric methods like OPTICS and Dirichlet GMMs (Ankerst et al., 1999; Görür and Edward Rasmussen, 2010), they had very poor or non-functional performance with the data, and a lack of troubleshooting support coupled with a desire to prioritize DBSCAN clustering led to these methods being abandoned for analysis. As was recommended in the Chapter 6, future projects will be developed in a Python IDE using TensorFlow functionality (TensorFlow, n.d.), which gives far more freedom in clustering models and more flexibility for parameter tuning.

7.6 Chapter Conclusion

This chapter has covered the exploration of an unsupervised clustering approach to walking activities for non-amputated and ILLA populations in free-living conditions. After studying the clustering behaviour and compromising on label resolution, an algorithm based on combining DBSCAN cluster models with a probability equation was created to identify the presence of stair movement, which could be clinically beneficial for monitoring the progress of ILLAs in the early stages of rehabilitation. Limitations of the model arose as a result of a low numbers for the study population, limited Matlab clustering options and inconsistency of tSNE cluster models. The next chapter is the final research-related chapter of the thesis, which looks at non-machine learning related analysis of walking activity, and an additional analysis of the ActivPAL's magnetometer in regards to how it could improve the supervised and unsupervised approaches discussed in this thesis.

Chapter Eight

Exploration of ActivPAL hardware and software for further development of activity monitoring capabilities for Individuals with Lower Limb Amputation

8.1 Introduction

The final research chapter of this thesis documents three independent experiments. Though independent, each of these investigations are thematically linked together as an exploration of the capabilities of the wearable device that was used in the study: the ActivPAL, as well as a validation of the VANE algorithm, the proprietary ActivPAL-specialised software created by the ActivPAL's manufacturer, PAL Technologies. In Chapter 3.2.3, when healthcare professionals were interviewed, some of the desired outcomes for an ILLA activity monitoring system included the walking speed, wear time and cadence. It was further stated in chapter Chapter 5.9.1 that due to the lack of reliable ground truth for walking speed, it would be impractical to validate this measurement. Instead, both cadence and walking speed can both be indirectly inferred through the step count, which is acquired through the VANE algorithm and validated by comparing with the video recordings. Validating the step count on lower limb amputees in free-living conditions is thus the first experiment.

Regarding the wear time of the ActivPAL (i.e when the subject affixed the ActivPAL on their thigh), it was also not possible to validate given that continuous video recording of the participants in their daily lives was unethical. In retrospect, the subjects could have been instructed to keep a log of when they attached and removed the device to provide some semblance of ground truth. As an alternative approach, the accuracy of standing detection by the ActivPAL's proprietary VANE algorithm was analysed in the second ActivPAL VANE experiment, again using the video recordings for ground truth. The implications of these findings and how a wear time algorithm could be constructed is discussed further in this analysis.

In Chapter 5.6.1, the magnetometer sensor of the ActivPAL was stated to be initially considered for inclusion but was discovered that the magnetometer signal values would occasionally saturate during a number of recordings. As discarding these recordings would lead to a significant loss in data, it was ultimately decided to abandon using the magnetometer in the primary investigation. The final ActivPAL analysis thus gives some insight as to how the introduction of magnetometer and magnetometer-based features would impact on both the supervised and unsupervised components of the investigation.

In the context of the overall research, this chapter demonstrates how the ActivPAL and the VANE algorithm can be used to further enhance the clinical activity monitoring system that has been discussed in this research.

8.2 ActivPAL Test #1: Validation of VANE on Lower Limb Amputee step count

8.2.1 Introduction

A major point of discussion in Chapter 2.7 was that the step count is a simplistic unit of measurement for activity, and by itself will not give an accurate profile of the ILLA's activity. However, by acquiring the step count, it is possible to calculate more informative measurements such as the subject's cadence and walking speed. Cadence can be acquired simply by dividing the step count over a period of time, and walking speed can be calculated by multiplying by the subject's average stride length, which can be acquired by an HCP in a clinical environment using rudimentary equipment (Dale, 2012; Bubnis, 2018; Perry and Burnfield, 2010).

$$Walking\ speed = \frac{Cadence * StrideLength}{2} \tag{8.1}$$

Axiomatically, so long as the step count is validated, then cadence and walking speed can also be validated, assuming a constant stride length. Stride length will change if the subject is traversing slopes or stairs (Lee and Chou, 2007; Sun et al., 1996), thus may require knowledge of the patient's stride length on slopes and stairs, coupled with integration of supervised HAR to adjust stride length accordingly. The ActivPAL experiments utilised ActivPAL's VANE algorithm for step detection. The VANE algorithm is proprietary, and its inner workings are not available for public knowledge, thus this experiment can only make general suggestions on how the algorithm could be improved. The research focused on validating the step count of the four ILLA subjects, given that they are the clinical group of interest of this research. The ActivPAL has received numerous validation studies over the years regarding its step count accuracy across a wide variety of demographics (Aminian and Hinckson, 2012; Cindy Ng, Jenkins, and Hill, 2012; Edwardson et al., 2017; Hergenroeder et al., 2018; Ryan, 2006; Storm, Heller, and Mazzà, 2015). Buis et al. (2014) and Pepin et al. (2018) both conducted studies on ILLA activity using an ActivPAL, but neither questioned the validity of the step count. The most comparable study is Deans et al. (2020), who validated the step count with an ILLA population in a controlled interior setting. Though the step count acquired high intra-class correlation coefficients, demonstrating agreeability between the ActivPAL and criterion's step counts, they concede that further validation is required in free-living settings. The study for the population also only included unilateral ILLAs. To the author's knowledge, this was at the time of writing, the first attempt to try and validate the ActivPAL's step count for ILLAs in a free-living environment, and potentially the first to validate the step count with a bilateral amputee.

8.2.2 Methodology

The data analysed for this experiment is the accelerometer and video reused from Chapter 5. To acquire the ground truth, the author watched each of the subjects' videos and manually counted the steps taken in the video with a mechanical counter. As the camera was fortunately angled downwards towards the subject's feet, step counts could be easily identified. Though a simple task, counting steps over long periods of time can be mentally taxing, and so to mitigate fatigue the researcher paused and tallied the step count at fifteen-minute intervals, taking a 1-minute break before resuming counting. To further mitigate counting fatigue, each step was technically counted as a stride, comprised of two steps, thus the count only increased when the same foot was moved forward. Due to ethical restrictions, the video data was only viewable by the author and their supervisors, and the supervisors at the time of carrying out the analysis were unavailable to partake as additional judges of the step count. As a compromise, the author performed the task three times on different days, with each separate day acting as a different judge of the step count. The calculations each day were performed in a different spreadsheet in Microsoft Excel™ (Microsoft, WA, USA) so that it would be difficult to remember the step count from the previous day and thus make the author pseudo-blind to their previous measurements. The procedure took inspiration from Deans et al. (2020) by only counting "purposive steps". In the context of Deans et al. (2020), purposive steps are those carried out by the participant to perform their assigned task. A purposive step is recontextualized in this study to define the steps taken to complete the walk. Deans et al. (2020) also measured "incidental" steps, which in their investigation are steps that are taken to traverse to the starting point of the exercise, and the steps taken to leave the room from the ending point of the exercise. These incidental steps will typically have different gait behaviour from typical forward-propelled walking and can include movements like shuffling and side-stepping. Purposive steps comprise the overwhelming majority of steps taken by the ILLA participants. In a free-living setting, incidental steps mostly arise from niche events such as the shuffling movement when opening and closing a gate, repositioning feet while waiting to cross a road, and backwards stepping. As found by Deans et al. (2020), the ActivPAL is considerably poorer at detecting incidental steps, and so when considering the general obscurity of these events in the dataset, it was decided not to count these incidental steps as part of the experiment.

Step counts as assessed via VANE are acquired in PAL Analysis' (PAL Technologies, Glasgow, UK) "Event" data simply by subtracting the cumulative step count corresponding to the end of the recording by the cumulative step count at the beginning of the recording. Step counts in VANE are technically stride counts (only count stepping data for a single leg), thus no further adjustments are required to compare the step count with the ground truth. There was some consideration as to whether the step counts should be tallied per each subject's individual recording, or the total count for each subject. By only looking at each subject's total step count, this provides some semblance as balance as each ILLA carried out roughly the same number of steps, but only provides 4 data points. Including each recording individually increases the number of data points from 4 to 12 but leads to imbalance in the number of steps counted due to half of the participants recording several times with shorter videos, while the other half recorded fewer videos but with longer recording times. A boxplot of the distribution of steps (as per ground truth validation) as seen in Fig. 8.1 demonstrates that though there is a larger amount of variance when including individual recordings, none of the recordings are statistically different enough in terms of step count to be considered outliers. It was decided to therefore perform validation on each individual recording, which also meant the results were more realistic, as a person would typically not carry out the same amount of walking activity either on a day-to-day basis.

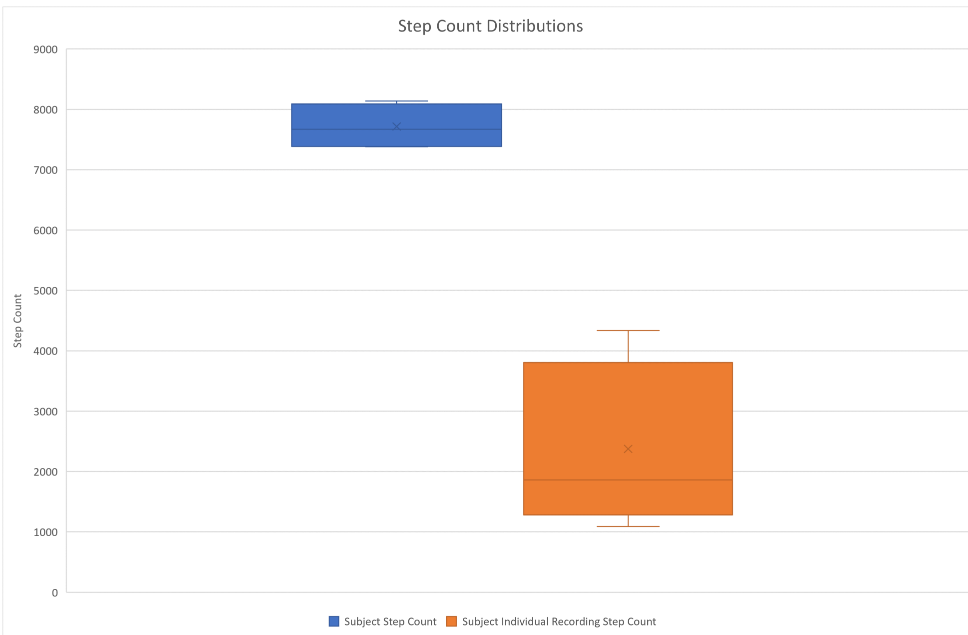


Figure 8.1 Box plot of step counts, comparing distributions of step count when accounting for all combined recordings of a participant, and when accounting for each individual recording of step count.

Agreement between the ActivPAL and criterion measurements of step counts was acquired through Intraclass Correlation Coefficients (ICCs) and was calculated in Matlab. All ICCs acquired use two-way models with random effects, using a single rater per measurement, or ICC (2,1) (McGraw and Wong, 1996; Shrout and Fleiss, 1979). The ICCs are first calculated using each of the three "ground truth" measurements to determine intra-rater reliability, then are applied again with the inclusion of VANE calculated steps. An ICC correlation coefficient of 0.7 or higher is considered to be the minimal acceptable level of agreeability, while a value of 0.9 or higher indicates strong reliability (Koo and Li, 2016). The compatibility of the VANE-calculated versus criterion steps are further compared using a pairwise t-test, setting a singular criterion value as the mean of the three ratings for each recording. The null hypothesis of a paired t-test states that the mean difference between observations is zero, thus a failure to reject the null hypothesis indicates further evidence of agreeability between VANE and the ground truth step counts.

8.2.3 Results

Intra-rater reliability in Table 8.1 demonstrated near perfect agreement (ICC = 0.99), this information is complemented by the raw step count data across the three days, where step counts are all very similar. When the VANE step count is introduced in Table 8.2, the ICCs again indicate near perfect agreement (ICC = 0.99), however the null hypothesis for significant differences between VANE and criterion count was rejected, indicating a systematic difference in the step counts. The raw data step count in Table 8.2 illustrates that, though the differences in step count are mostly minimal, there is still a statistically significant systematic decrease in the number of steps between the VANE count and the criterion count.

Table 8.1 Intra-rater reliability of step count in ILLA recording sessions

Subject/ Recording Iteration	Day 1	Day 2	Day 3	Mean Criterion Step Count	VANE Step Count	Step Count Difference
A1-1	1456	1458	1452	1455 ± 3	1434	21
A1-2	1430	1427	1441	1433 ± 7	1422	11
A1-3	1471	1473	1473	1472 ± 1	1453	19
A1-4	3024	3023	3030	3026 ± 4	2973	53
A2-1	3802	3835	3845	3827 ± 23	3688	139
A2-2	4334	4350	4360	4348 ± 13	4061	287
A3-1	3812	3791	3820	3808 ± 15	3703	105
A3-2	4140	4129	4160	4143 ± 16	4072	71
A4-1	1089	1090	1074	1084 ± 9	1018	66
A4-2	1860	1867	1863	1863 ± 4	1837	26
A4-3	2201	2212	2216	2210 ± 8	2180	30
A4-4	1107	1115	1119	1114 ± 6	1090	24
A4-5	1134	1147	1139	1140 ± 7	1116	24

Table 8.2 ICC analysis of VANE and criterion step counts, with additional pairwise t-test.

Intra-rater ICC	UB	LB	VANE-Criterion ICC	UB	LB	Pairwise t-test p value
0.99	1	0.99	0.99	0.99	0.99	**0.0078**†

UB = Upper Bound, LB = Lower Bound, † = statistically significant

Notably, the error in step count is significantly higher in subject A2 compared to the other subjects. This discrepancy in error can be attributed to two factors: subject A2 had their camera angled at a higher position than others, such that during periods of the recording their feet were not visible, and step count was instead judged by the swing of their arms, which also resulted in high intra-rater standard deviations due to a change in counting methodology. The other discrepancy of the step count appears to arise during the beginning of their recordings, where they ascend and descend a flight of stairs several times. By plotting the length of the step count as detected by VANE in Fig. 8.2, it can be seen that throughout the exercise the step count lengths exceeded 1.5 seconds, indicating VANE has missed a step. This highlights a weakness in the VANE's algorithm for detecting step movement on certain configurations of stairs.

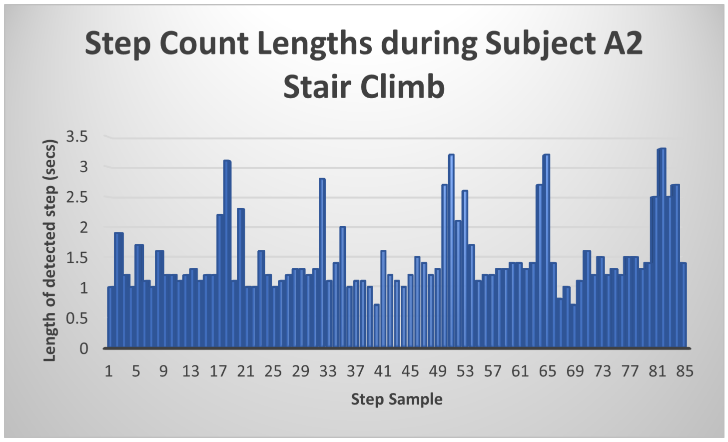


Figure 8.2 Bar chart of the durations of detected steps during one stair climbing session performed by participant A2

8.2.4 Discussion

The results of this study implicate that the VANE algorithm, though able to achieve comparable step counts to the ground truth, systematically underreported the step count of ILLAs (including a bilateral amputee) in free-living conditions. This finding aligns with the research carried out by Deans et al. (2020) for their purposive step measurements. Despite the much smaller step sample sizes (mean 106 steps vs. 2379 in this investigation), the ICCs and p-value for paired t-test were comparable (ICC = 0.93, p = <0.0001). The increase in ICC and p-value in this investigation likely stem from having much higher step counts, diminishing the proportion of the error in the step count. Though the set-up of the experiments is different, the error in step count from the VANE algorithm likely stems from the same origin: a decrease in walking cadence. In Deans et al. (2020) , participants who had a slow walking cadence typically had less accurate reported step counts compared to the ground truth. Similarly, Larkin et al. (2016) also found a systematic underreporting of the VANE step count in patients with rheumatoid arthritis under laboratory conditions, who would also have a lower cadence than the healthy population. In this investigation, the ILLA volunteers either had many years of experience as an amputee ambulating with a prosthesis or were relatively fit and active. This reflects in their average cadences, listed in Table 8.3, which was estimated by taking ten random samples of video in each subject where ambulation is uninterrupted and manually counting the number of steps. These cadences are comparable to middle-aged adults with no known gait morbidities (Tudor-Locke et al., 2020), suggesting these are above average cadences for a typical ILLA. Due to their high cadences, the VANE algorithm has little issue in detecting their steps, with the exception being during stair movement where cadence naturally decreases to acquire firm footing on each step. The VANE algorithm could be potentially improved in future iterations by fusing information from the machine learning classifiers discussed in this thesis: in the detected presence of stairs, the VANE algorithm could adjust their step detection refractory period (a known component of the VANE algorithm which PAL Technologies have granted permission to acknowledge) to account for the slower cadence.

Table 8.3 Cadence of ILLA Participants

Subject	A1	A2	A3	A4
Average Cadence (Steps/Min)	110	120	122	116

As the study was not designed around validating the step count, the methodology quality of this study is comparatively weak in regard to Deans et al. (2020) . As mentioned in the results, there were occasions in the recorded videos where feet movement was not visible in the camera. Deans et al. (2020) on the other hand used a multi-camera set-up and assessed each step count through the median value of all camera positions, giving a much more accurate assessment of the ground-truth. While Deans et al. (2020) states that using separate judges was not necessary for their project, having independent judges in this study would have given a better unbiased estimate of the step count. Future investigations to validate the step count in free-living conditions would require much higher sampling sizes with more diversity in unilateral and bilateral amputees, as well as including transfemoral amputees who may exhibit different behaviour due to requiring the ActivPAL to be attached to the prosthesis instead of a sound thigh. Attaching a second camera, or a camera with a wider field of view in the vertical direction, would also ensure that the criterion measurements match with the absolute true step count. Deans et al. (2020) found that there were no significant differences in ActivPAL step count when a separate ActivPAL was attached to the sound thigh, but nonetheless would warrant further replication in a free-living setting. Some further testing in natural interior environments (e.g the subject's home or work) would also give a greater indication of the VANE step count in true free-living conditions.

8.3 ActivPAL Test #2: Validation of VANE on ILLA standing detection

8.3.1 Introduction

The second ActivPAL experiment was an assessment of the VANE algorithm's detection for the presence of standing time. Like step count, the accuracy of the ActivPAL's standing behaviour validation has been a frequently researched topic and is in fact used as the criterion standard for standing detection in some investigations (Carpenter, Yang, and West, 2021; Dowd, Harrington, and Donnelly, 2012; Hamer et al., 2020; Larkin et al., 2016; O'Brien et al., 2020). Deans et al. (2020) carried out analysis of the VANE algorithm with a ILLA population, however they did not directly measure standing times, only sedentary (sitting/reclining) time. This should therefore, at the time of writing, be the first study to attempt to validate the ActivPAL's standing behaviour for ILLAs under free-living conditions.

8.3.2 Methodology

The data utilised in this study differs from that used in the previous ActivPAL experiment. Due to the nature of the study, most of the recordings of the ILLA participants involved a largely uninterrupted walk from start to finish, and standing events were very infrequent. The exception to these general observations is subject A3 and two of the recordings of subject A4. In subject A3's scenario, during both of their recordings they sporadically halt their walk for various reasons such as to check their phone or the status of the recording camera's battery. Subject A4 went on a walk with their dog in two of the recordings (recordings 2 and 5), which led to frequent interruptions in the walk due to the dog pulling on its leash to investigate its surroundings. These four recordings are the only recordings of the ILLA participants to provide ample scope for the detection of standing activity.

The ground truth for analysing stepping behaviour is once again acquired through manual observation of the recorded videos. Stoppages of a length of three seconds or greater were manually noted down in a spreadsheet, three seconds was heuristically chosen as the threshold to mitigate recording of brief pauses during gait, for example if the participant readjusts their feet before ascending or descending a flight of stairs. Due to the overall small number of stoppage events, and the ability to go back and rewind footage, only one set of criterion measurements was needed for this experiment. The standing times are then extrapolated into a vector, with the total length of the vector equal to the length of the recording in seconds. Standing and moving times from the VANE algorithm are acquired through the "Event" .csv file generated in PAL Analysis. A Matlab script read the .csv file and extracted the event type for every second of the recording. An event type of "1" indicates standing, while an event type of "2" indicates walking. Thus, the ground truth and VANE analysis both generate two vectors of equal length. Due to using categorical data, the "agreement" between the criterion and the VANE algorithm (otherwise known as the interrater reliability) are analysed using Cohen's kappa coefficient (Cohen, 1960). The smallest sample size on a per recording basis is 2,272 samples, which indicates the sample size requirements for instigating the coefficient are sufficient based on the assumption that the marginal rating frequencies between the criterion and VANE measures are the same (Bujang and Baharum, 2017). As recommended by McHugh (2012), a kappa coefficient of 0.6 or less indicates weak agreement and that only 25% of the data can be considered reliable. A kappa coefficient of 0.8 or greater indicates strong agreement and up to about 80% of the data can be considered reliable. The calculated kappa coefficients are supplemented with precision, recall and F1 scores of both activities as derived from a confusion matrix.

8.3.3 Results

The precision, recall, F1, and kappa coefficients are detailed in Table 8.4. The precision and recall of walking events remain consistent throughout each of the 4 recordings. Any decrease in the precision values for walking can mostly be explained by synchronizing errors between the criterion and VANE's detection of standing behaviour, for example recognizing the start of leg movement half a second before the VANE algorithm detects the presence of walking. VANE however appears to struggle at recognizing the presence of standing behaviour when transitioning from walking movement. Included in Fig. 8.3 are stair plots of the activities as represented by both the ground truth and VANE, where an agreement of activity is indicated by green lines. In the stair plots for subject A4, it can be seen that there are multiple instances of standing activity occurring in the ground truth, while VANE does not detect the presence of stationary movement for the entirety of the event. When comparing the timestamps to the video, these situations arise when the subject is forced to quickly halt their walking movement, for instance turning around to look at their dog. The duration of these standing activities is short enough that a predetermined threshold in VANE's algorithm does not correctly recognize this duration as the subject standing.

Table 8.4 Analysis of standing and walking time precision measured by VANE versus ground truth observation

Recording	A3-1	A3-2	A4-2	A4-5
Precision (standing)	0.99	0.91	0.96	0.93
Precision (walking)	0.98	0.99	0.93	0.92
Recall (standing)	0.77	0.76	0.61	0.51
Recall (walking)	0.99	0.99	0.99	0.99
F1 (standing)	0.86	0.82	0.74	0.65
F1 (walking)	0.99	0.99	0.96	0.96
Cohen's Kappa	0.85	0.81	0.71	0.62

8.3.4 Discussion

The main interpretation of this analysis is that intentional stopping in walking activity is more easily recognizable for the VANE algorithm than sporadic halts in walking activity, as reflected by the kappa scores for subjects A3, who tended towards intentional stopping, and A4, whose dog monitoring created frequent events of sporadic stopping. After discussion with PAL Technologies, it was revealed that in VANE detection, the maximum allowable step count length is six seconds. Therefore, any brief standing event shorter than this length of time will be regarded as a step regardless of how "still" the standing behaviour is. This explains why the kappa values for subject A4 are considerably lower than subject A3, despite being carried out over a much shorter period of recording time (70/75 minutes for A3 vs. 40/25 minutes for subject A4). In comparison with similar research, Bourke, Ihlen, and Helbostad (2019) who carried out validation of the ActivPAL for non-amputated participants in both laboratory and free-living settings found standing detection was actually recognized better in a free-living scenario and acquired comparable F1 scores to this investigation (73.5%

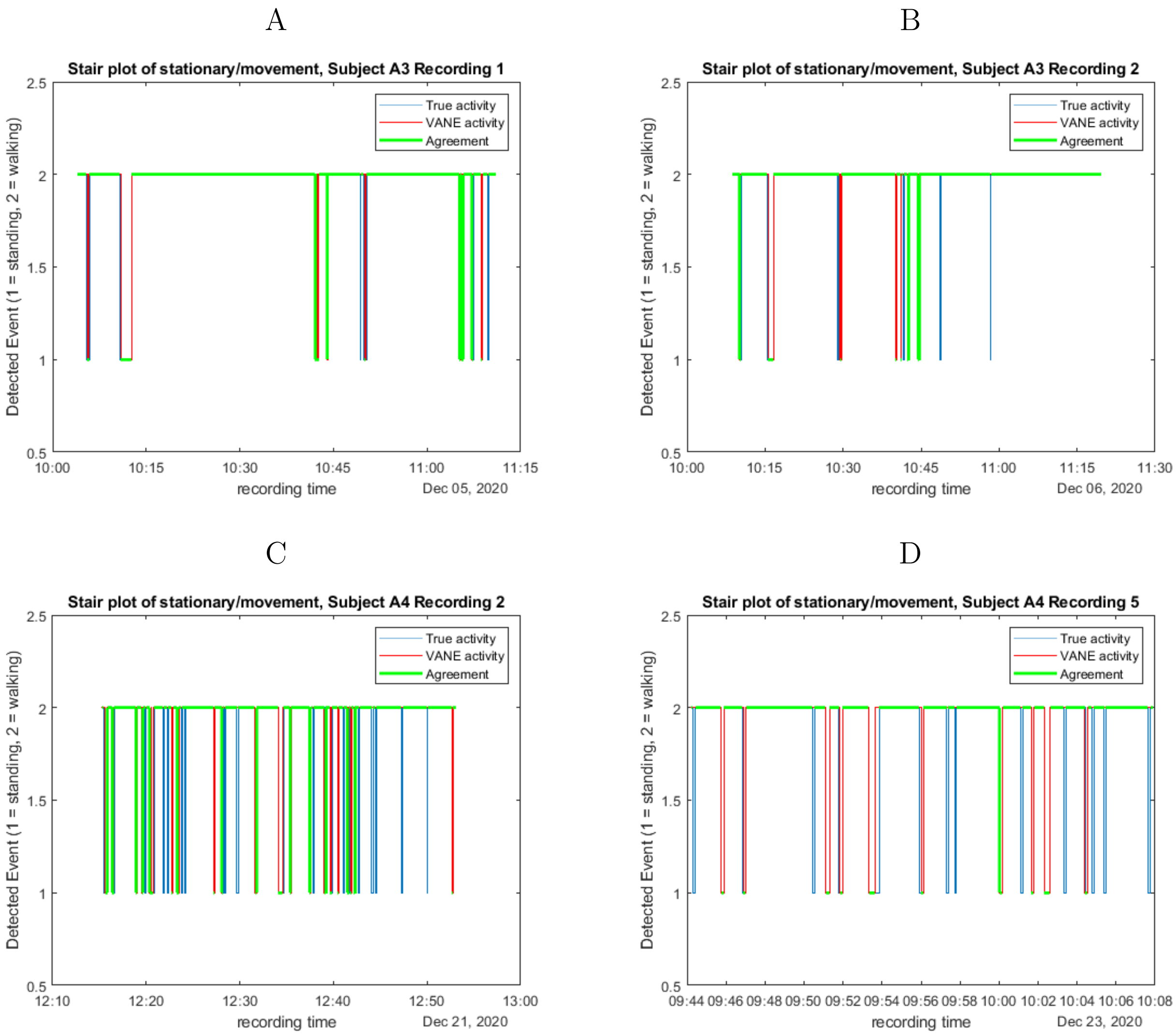


Figure 8.3 A: Stair plot of standing and stationary movement for the 1st recording of participant A3 || B: Stair plot of standing and stationary movement for the 2nd recording of participant A3 || C: Stair plot of standing and stationary movement for the 2nd recording of participant A4 || D: Stair plot of standing and stationary movement for the 5th recording of participant A4

F1 score for lab vs. 76.4% for free living). Their Cohen's Kappa was also comparable to those acquired by subject A3 (0.86) while also including sitting detection in their measured activities. A major difference however was that the free-living component was carried out in the subject's home rather than outdoors, suggesting that the findings from this investigation hold true in interior environments as well as outdoor environments.

While it is unlikely in a clinical context that an ILLA would choose to go on a walk with a dog, especially in the early stages of rehabilitation, sporadic standing events could arise in other scenarios. Cooking, cleaning, doing laundry, and shopping are other exemplary events where the subject will be in a continuous cycle of standing for brief intervals followed by moving. If the results from subject A4 reflect on VANE's general ability to recognize sporadic movement, then in these scenarios the VANE algorithm may interpret these movements as constant motion and thus the VANE algorithm requires further development of their standing detection.

The small sample size means these findings may not generalize to the ILLA population and as such should be treated as a case study rather than a general observation. Future studies with a similar research aim should encourage its participants to take regular standing and sitting breaks during their regular walks to acquire a more realistic idea of how VANE can handle the recognition of these standing bouts. The validation of stationary times via video observation should also be carried out by multiple independent researchers to avoid introducing observer bias.

8.3.4.1 Implications for Presence of "Null Activity" and Wear Time Detection

When annotating the activities for the recordings in Chapter 5.5 it was noted that there was a required presence of the "null" class to try and remove potentially confusing sections of the recording from the machine learning process. These "null" instances were occasionally comprised of standing behaviour. If the machine learning algorithm were to be tested on new data that had not been annotated and pre-processed to remove unwanted data, this would introduce confusion into the machine learning algorithms, whether approaching the problem from a supervised or unsupervised perspective. These scenarios are where the VANE's standing detection reliability (as shown in Grant et al. (2006)) can provide additional assistance in removing these blocks of activity from the testing data. Referring back to Table 8.4, it is notable that the precision of standing recognition is much higher than the recall, indicating that there is a lack of false positive recognitions of standing behaviour. In the context of activity recognition, this means that VANE algorithm will not detect false presences of standing movement during gait and will therefore not accidentally remove walking data from recognition. Conversely however, the recall of standing was considerably lower, which would implicate some standing activities into the test set of data. A potential solution to the weaker recall would be to additionally analyse the duration of the stepping event as calculated by VANE, removing any instances of a step movement longer than a pre-determined threshold (e.g 4 seconds), however the drawback to this approach would be that certain stair or slow walking movements could be removed from data analysis due to the decreased cadence. Perhaps a more elegant solution would be to perform pre-clustering of the feature data before activity recognition process to identify and remove standing activities, which tend to form distinct clusters in low dimensional spaces (Abedin et al., 2020; Ma et al., 2021). Automated standing detection warrants further investigation in future studies.

The presence of standing activity via VANE analysis can be further extended to analyse the wear time of the ActivPAL device. Though unfortunately unverifiable within the scope of the investigation, some research has been dedicated to the development of wear time detection of the ActivPAL (Carlson et al., 2021; Winkler et al., 2016). While wearing detection can easily be identifiable for instance through non-sporadic periods of walking and standing activity, distinguishing between non-wear and prolonged bouts of "sitting" activity (e.g sleep) can be considerably more challenging. Fortunately, with the introduction of a magnetometer in the latest models of ActivPAL, identifying the orientation of the ActivPAL device is now a possible feat; Fig. 8.4 shows an example of the orientation data as illustrated in PAL analysis for one of the subjects included in this study. Using rudimentary algorithms, non-wear could be easily recognized by abnormal orientations of the ActivPAL (i.e left, right, front, or down orientation) over extended periods of time. To the author's knowledge, the orientation information of the ActivPAL has yet to be validated and so warrants further testing in future studies.

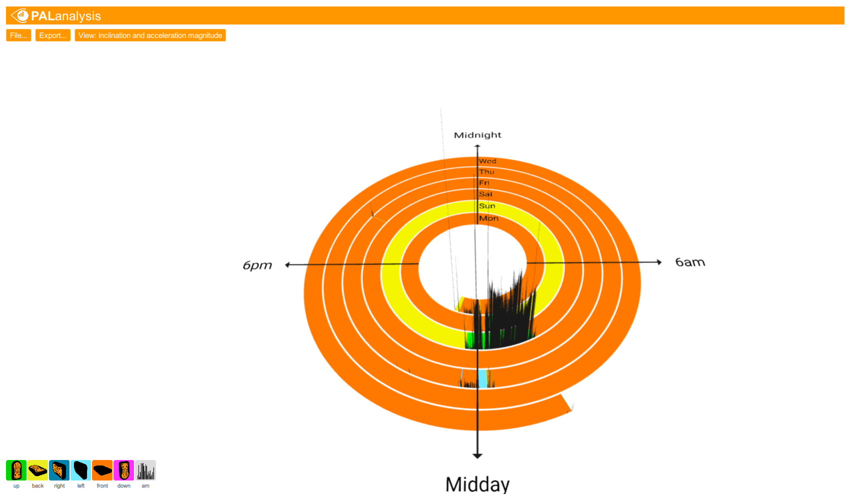


Figure 8.4 Spiral graph of ActivPAL orientation for one of the participants during their entire recording duration

8.4 ActivPAL Test #3: Inclusion of Magnetometer features

8.4.1 Introduction

The final post hoc experiment is a brief venture into the magnetometer functionality of the ActivPAL and its impact on the supervised and unsupervised components of this thesis. As explained in Chapter 5.6.1, the magnetometer data ultimately had to be discarded from the main analysis. This experiment reintroduces the magnetometer and magnetometer features, exclusively looking at subjects where the magnetometer did not saturate. Magnetometers in the context of how they can benefit (and hinder) HAR studies was explained in detail in Chapter 4.2.2.

8.4.2 Methodology

The number of viable subjects was limited to four participants: Non-amputated subject H8, and ILLA subjects A1, A2 and A3. All other subjects experienced periods of magnetometer signal saturation, and so were considered unusable. The magnetometer signals are pre-processed using the same configuration as the accelerometer; that being a Butterworth bandpass filter with passband frequency ranges of 0.3 – 4Hz. The resulting signal has a smooth sinusoidal waveform that requires no further denoising, as shown in Fig. 8.5. The magnetometer signal is subsequently segmented with non-overlapping sampling windows of 40 samples in length.

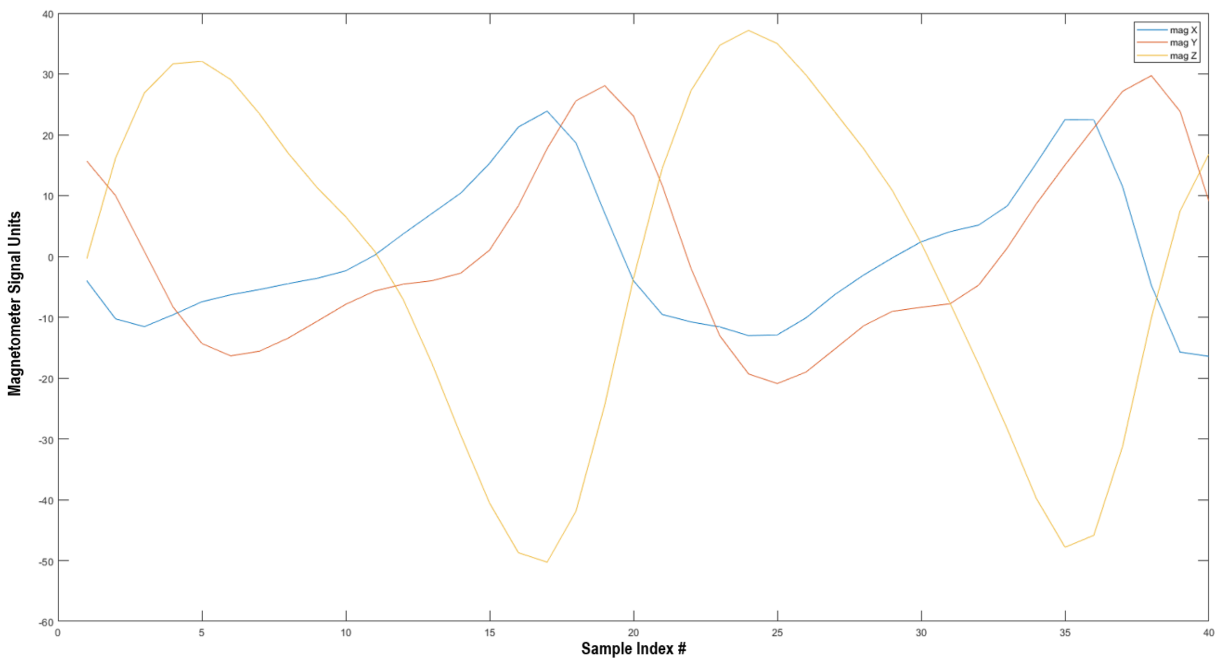


Figure 8.5 Random sample of timeseries filtered triaxial magnetometer signal

The magnetometer features calculated in the study are identical to those calculated with the accelerometer, in order to determine whether the same feature calculations would improve the machine learning performance when including the same features applied to the magnetometer. In other words, to determine whether classifier efficacy could be improved through simple implementation of a magnetometer and not through the addition of magnetometer specific features. An additional 9 features are created from the cross-correlation coefficients of the XYZ axes of the accelerometer and magnetometer, resulting in a total of 485 features. Features undergo the same standardization process as described in Chapter 5.8.1, resulting in zero mean and a variance of one for each feature.

The supervised component exclusively focused on how the recognition performance of an SVM classifier changed with the addition of magnetometer features. The dataset combines data from all 4 individuals and used Level 1 of label resolution (level walking, uphill, downhill, upstairs, downstairs). The dataset dimensionality is reduced via PCA to conserve 95% of the total systematic variance, resulting in approximately 110 principal components when magnetometer features are included, and 60 components when only accelerometer features are used. Due to the significant reduction in size of dataset, it was considered unfeasible to explore the previously employed LSTM network. The SVM classifier undergoes the nested cross validation process previously described in 6.2.5. The data relegated to the inner loop is balanced with SMOTE and then used to tune the SVM classifier for optimal hyperparameters (tuning for Kernel type and Box constraint), the resulting trained classifier is then validated with the testing data partition, the resulting recognition accuracies are then obtained as a result of 5-fold cross-validation.

For the unsupervised component, the data is separated by each model. tSNE, with the same hyperparameter configurations as in Chapter 7.2.6, reduced the dataset from its original size to two dimensions. The analysis of the unsupervised component extends only to an analytical look at the clustering quality of the model when labelled for the base level of resolution. The algorithm developed in Chapter 7.3.2 is not used.

8.4.3 Results

The introduction of magnetometer features led to a significant performance boost in recognition accuracy, achieving a mean recognition accuracy of 87%, compared to using only the accelerometer features which had a mean recognition accuracy of 73.8%. The accuracy of the accelerometer features fell from 77.2% in Chapter 6.3.2, which is expected given the overall smaller training dataset. This could also indicate that had the magnetometer been fully functional throughout the investigation, the mean recognition accuracy could have exceeded

90%. The resultant confusion matrices for both approaches are presented in Figs. 8.6-8.7 A notable weakness of including magnetometer features is that despite having good recall values for all activities, the precision of uphill and downhill activities remains poor (63.6% for uphill, 58.4% for downhill).

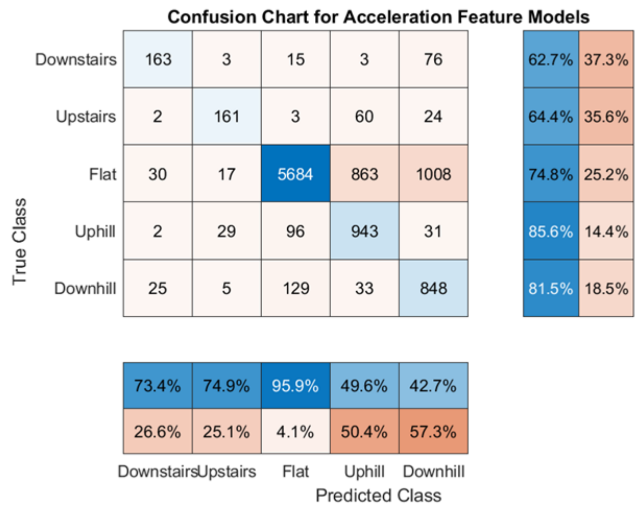


Figure 8.6 Confusion chart of SVM classifier at level 1 label resolution using only accelerometer features.

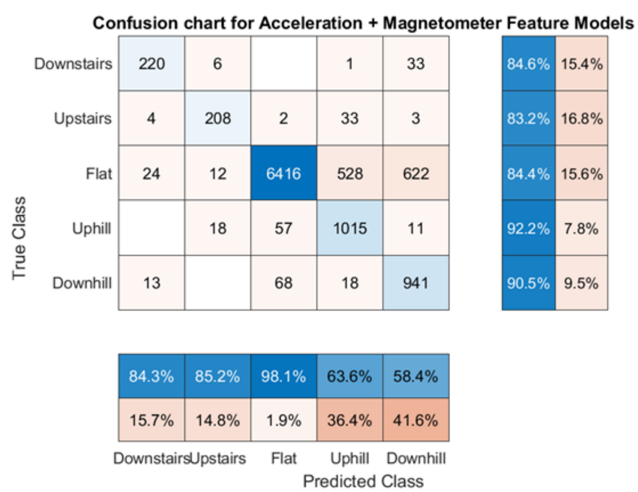


Figure 8.7 Confusion chart of SVM classifier at level 1 label resolution using combined accelerometer and magnetometer features.

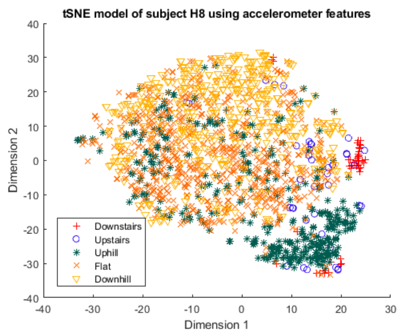
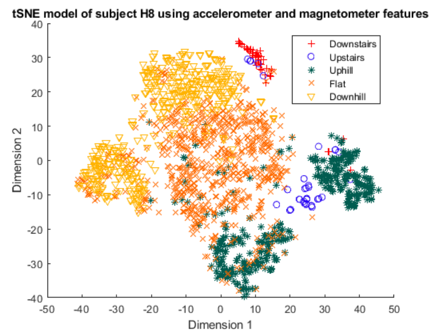
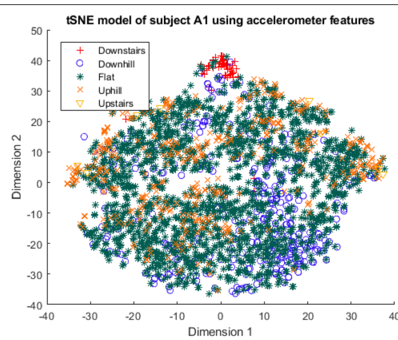
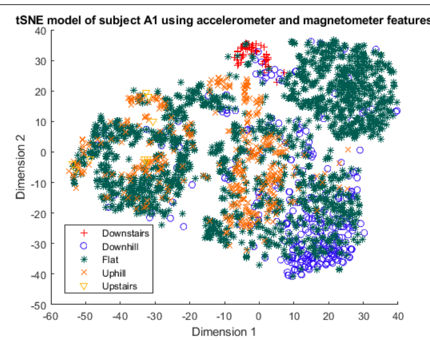
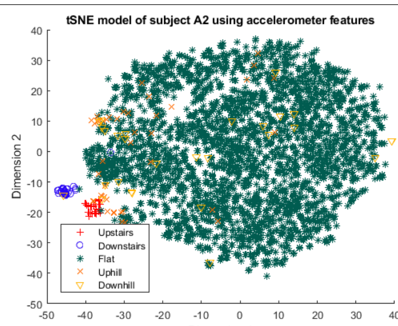
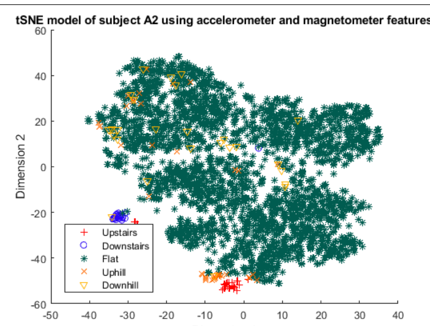
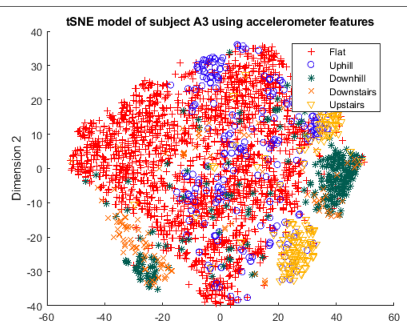
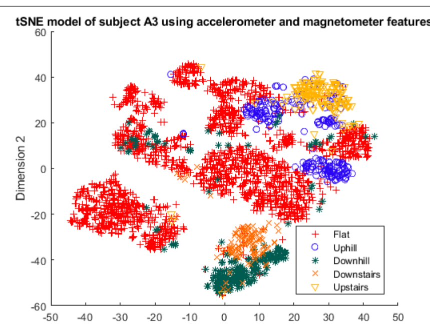
The resulting tSNE clustering models when magnetometer features were included proved to have more powerful clustering effects than when only using accelerometer features. As

Table 8.5 demonstrates, not only is the model able to separate upstairs and downstairs data, in subjects H8, A2 and A3, there is an appreciable formation of hill clusters which previously proved impossible to form in Chapter 7. Subject A1 was the only participant where the inclusion of magnetometer features did not lead to an appreciable visual improvement in the clustering model, though at the least there is a stronger concentration of downhill movement in a single area compared to the accelerometer only model.

8.4.4 Discussion

Contrary to the literature review of magnetometer features in Chapter 4.2.2, the addition of magnetometer features had a marked improvement over the use of only accelerometer features for the supervised approach. For example, the inclusion of magnetometer data in Garcia-Gonzalez et al. (2020) only led to a 2% increase in recognition accuracy (60.1% to 62.6%), while in this experiment there was an improvement of 13%. An important methodological difference is that they include GPS readings with the accelerometer readings in the initial feature matrix calculation, indicating that GPS-acquired features could have overlap with magnetometer-acquired features in terms of their similarities. The improvement of classification in San Buenaventura and Tiglao (2017) was also only 0.6% greater than using only an accelerometer, though due to a high base recognition accuracy (96.9% with accelerometer only), there was little room for improvement. The takeaway interpretation from these findings is that, when the base recognition accuracy of the measuring device is low, it is worthwhile exploring the introduction features from other sensors. On the other hand, while this approach to magnetometer feature extraction worked in favour of the dataset in this experiment, these results may not generalize on a larger population; Shoaib et al. (2014) found in their HAR investigation that random combinations of sensors did not see an improvement in recognition performance, and rightfully argues that the features of each sensor should be carefully considered in terms of their relevance.

Table 8.5 Cluster models comparisons with and without the inclusion of magnetometer features

Subject	Accelerometer features only tSNE models	Accelerometer & Magnetometer features tSNE models
H8		
A1		
A2		
A3		

Judging from Ariza Colpas et al. (2020)'s systematic review of unsupervised human activity recognition and the author's own research, there is a comparative lack of investigations that have directly compared magnetometer and accelerometer features in the context of cluster analysis. At the time of writing, seemingly all unsupervised approaches with wearable sensors have either exclusively used accelerometer sensors, or at best combined both sensors without a direct investigation on the effect of magnetometer inclusion. While the visual clustering results obtained in this investigation show promise, the proposed algorithm from Chapter 7.3.2 would require serious modifications to adapt to the new clustering behaviour shown. Particularly, one of the main evident changes is that there is no longer a centralized walking cluster, instead there are scattered smaller walking clusters. One of the likely important factors driving the cluster separation are changes in the ActivPAL's orientation. As shown in Fig. 8.8, by re-plotting the cluster models and labelling the data according to the recording in which they were carried out, it can be seen that especially in subjects A2 and A3, there is some distinctive separation of the data between the different recordings. Given that there are only four subjects, half of which experienced this phenomenon, this would require testing with a larger population to determine whether orientation changes were truly an important factor in clustering behaviour.

Though the inclusion of the magnetometer led to a considerable improvement in the performance of the supervised and unsupervised approaches, it was unfortunate that the magnetometer-based results were only valid for a third of the participants. While there are actions that could be taken to try and prevent the magnetometer from being decalibrated, most actions will be difficult to implement consistently within a clinical setting. PAL Technologies are working towards a magnetometer calibration algorithm, however as of the time of writing (August 2021) there are currently no indications that the algorithm is complete and would still require validation after completion. The introduction of a user-led calibration procedure, even for something as simple as placing the ActivPAL on its different sides for several seconds, can lead to a marked increase in device compliance requirements. However,

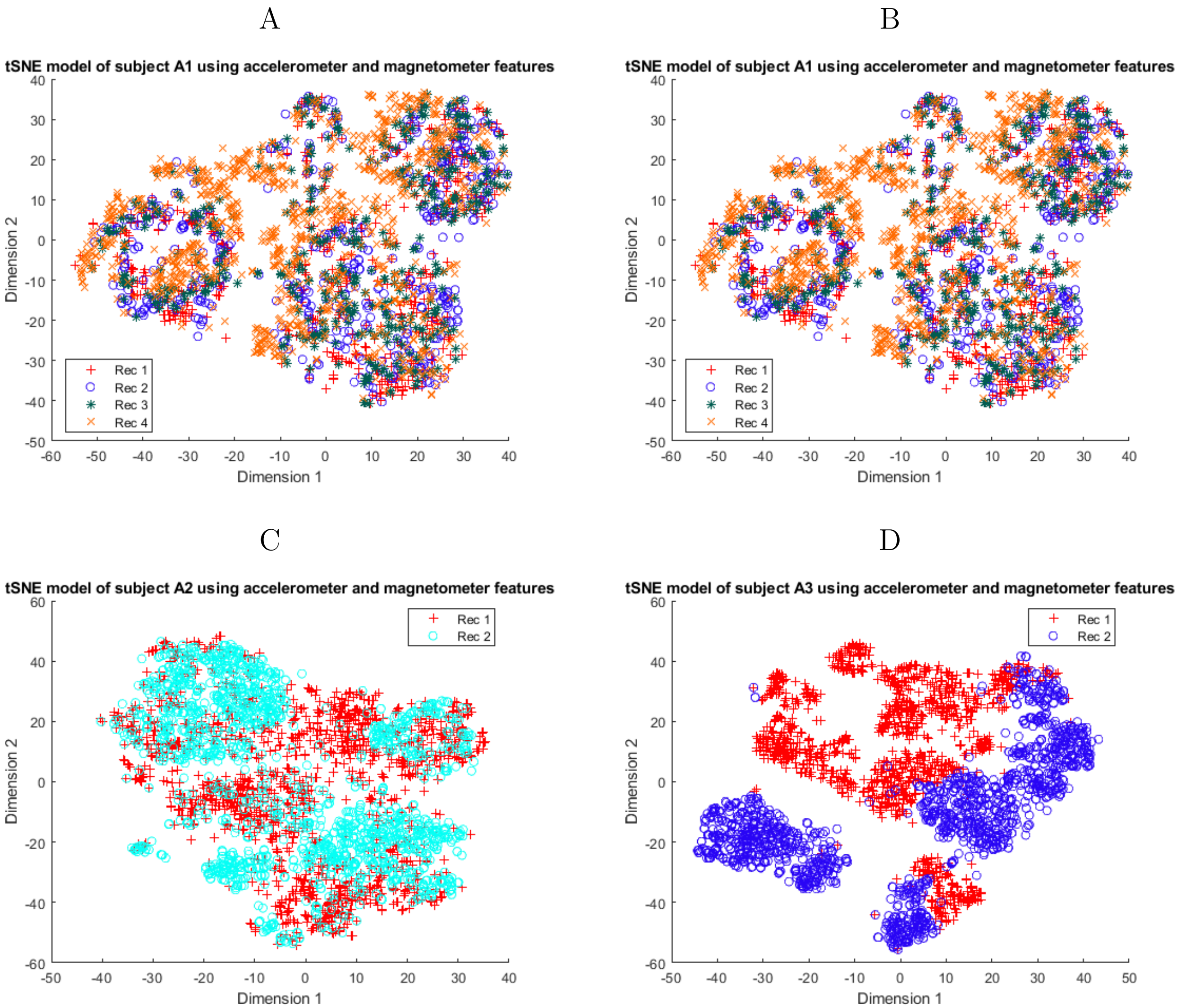


Figure 8.8 A: Cluster plot of subject H8 using tSNE dimension reduction on the combined accelerometer and magnetometer feature extraction matrix, labelled by each separate recording. || B: Cluster plot of subject A1 using tSNE dimension reduction on the combined accelerometer and magnetometer feature extraction matrix, labelled by each separate recording. || C: Cluster plot of subject A2 using tSNE dimension reduction on the combined accelerometer and magnetometer feature extraction matrix, labelled by each separate recording. || D: Cluster plot of subject A3 using tSNE dimension reduction on the combined accelerometer and magnetometer feature extraction matrix, labelled by each separate recording.

there is no guarantee that the user will comply with the calibration procedure and correctly carry it out. Likewise, magnetic disturbance could potentially be mitigated through the introduction of Kalman filter (Roetenberg et al., 2005), but this would require an additional gyroscope component and would increase the production cost of the ActivPAL. Additional research would also be necessary to ensure that any interference filtering is also not removing useful information (i.e leg movement) from the magnetometer signal.

8.5 Chapter Conclusion

The three experiments conducted in this chapter provided useful insight regarding the quality of the ActivPAL device in a clinical activity monitoring context. For an ILLA population in free-living conditions, it was found that the VANE algorithm was reliable in acquiring accurate step counts and detecting purposive standing events mid-walking. However, a drawback of the VANE algorithm was illustrated by its tendency to underreport step count when there is a decrease in cadence, most notably during stair movements, and was insensitive for detecting sporadic stationary events. Nonetheless, standing detection by the VANE algorithm could be a powerful tool for automatically detecting and removing "null" activities from testing data, leading to respectable recognition performances and strong clustering models. The inclusion of features from the ActivPAL's magnetometer also led to a significant improvement in both supervised recognition accuracies, as well as more powerful clustering effects, particularly in uphill and downhill movement. But, given that the magnetometer only operated correctly in only four of twelve participants, some substantial research efforts will be required to ensure correct calibration and remove the presence of magnetic interference without increasing user compliance or removing important data from the signal. Taking all three experiments into account, the ActivPAL can enhance future iterations of the activity monitoring system by giving detailed breakdowns of standing and stepping activities of an ILLA, provide methods of detecting wear time and cadence, and use the magnetometer to further improve the recognition accuracies and clustering capabilities of supervised and unsupervised machine learning systems respectively, potentially expanding on the scope of walking activities that can be recognized.

Chapter Nine

Conclusion: Implications of Research in Clinical Context & Retrospective Evaluation

The final chapter of the thesis is comprised of two main components. The first component describes how the research carried out in this thesis can be used to construct the framework of a clinical-based activity monitoring system. This component is structured in a Frequently Asked Questions (FAQ)-type format, such that the main implications of the framework are clear and concise. The second component is a retrospective analysis of the work carried out in this thesis, and an outlook of what work should be undertaken in the future.

9.1 Implications of a Clinical Activity Monitoring Framework

9.1.1 What is the clinical activity monitoring framework?

The clinical activity monitoring system framework is comprised of a central activity monitoring system which acquires physical IMU data from an ActivPAL device worn by a user and interprets the data via machine learning algorithms to output quantitative physical activity information. The rest of the framework details protocols and recommendations for the system, which will be explained in the oncoming segments. Fig. 9.1 represents a summary of the "loop" in the clinical activity monitoring framework, wherein a HCP gives an ActivPAL to the ILLA, they perform physical activity over the course of a week, and return the device to the HCP. The HCP connects the ActivPAL to a PC where an application reads and interprets the data from the ActivPAL and outputs the key activity monitoring outcomes in a Graphical User Interface (GUI). The HCP will interpret information relayed by the GUI to determine what actions should be taken to help improve the ILLA's situation. The process can then repeat for as many times as needed.

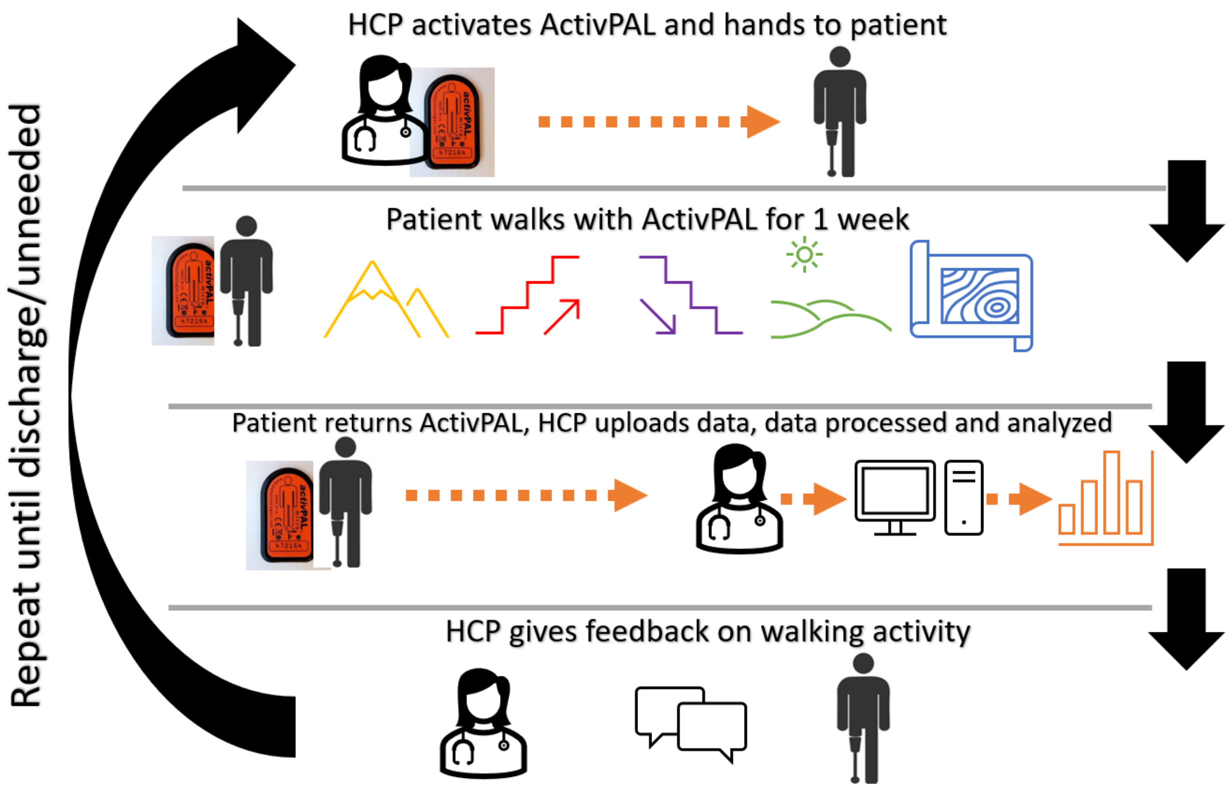


Figure 9.1 High-level infographic of the clinical activity monitoring framework

9.1.2 Who is the target demographic?

The framework is primarily targeted at HCPs that specialize in the rehabilitation of ILLAs. For example, physiotherapists whose general goal is to try and promote or improve their client's physical activity would benefit the most from this system. From the findings of Chapter 3.3.1, the evidence suggests that framework has maximum efficacy at the beginning stages of prosthetic rehabilitation, when the amputee is first fitted with a prosthesis. In these stages, the amputee is more likely to comply with the recommendations set out by the HCP if they believe that following the HCP's advice is vital to their rehabilitation progress, and so are more likely to consent to being outfitted with an ActivPAL and permit their activity to be analysed. The ActivPAL should be suited for deployment particularly within rehabilitation clinics or hospital wards, and also suitable for short term (week-long) monitoring in a home environment, as the evidence from this thesis suggests.

Beyond the early stages of prosthetic rehabilitation, there are diminishing reasons why the clinical activity monitoring system should be used in favour of using consumer grade fitness trackers such as a Fitbit®. The primary reasoning is that, from Chapter 3.3.1, it was found that all interviewed ILLAs, who had experience of prosthetic rehabilitation of a minimum of one year, no longer maintained a relationship with their healthcare provider (e.g NHS staff) beyond emergency situations relating to their prosthesis. The interviewed HCPs also conceded that contact frequency becomes minimal past the initial stages of rehabilitation (hosting check-up meetings every few months or years). Considering that an ActivPAL unit has battery and memory capacity suited for a week, the ActivPAL in its current iteration has no practical use for a clinical monitoring system when the average contact frequency between amputee and HCP is greater than two weeks. In order to make the ActivPAL viable for long term monitoring, this would require the development of a cloud server from which ActivPAL data could be remotely stored and retrieved by the HCP, such that the data on the ActivPAL hardware can be safely overwritten without data loss. The development of such a framework is beyond the scope of the thesis and the author's expertise, but even if such a framework were to exist, this would incur additional compliance from the ILLA user, as they would need to initiate transfer of the data from the ActivPAL either to a computer or a mobile app via Bluetooth communication. There would also be additional financial costs for server hosting, and there would be some privacy concerns regarding who can access the data if it was being held in cloud storage. Going further, even if theoretically there were no financial, privacy or compliance concerns, an ILLA client may simply not be interested in having their activity monitored and judged by an individual, especially on a long-term basis. If they were still interested in monitoring and upkeeping their physical activity, a far simpler option would be to purchase a Fitbit® for £50 (Fitbit, n.d.[b]), or a pedometer for £25 [1]. Though these devices may not be able to capture the depth of data acquired by the ActivPAL in a clinical activity monitoring framework, they can still acquire relevant health-based information for a fraction of the price (see Section 9.1.7) and with minimal concerns for privacy.

9.1.3 How is data acquired?

To analyse the activity of an ILLA, the HCP should provide their client with an ActivPAL device, the hardware revision should be a 3 micro at minimum to be compatible with the latest PAL software (PAL-Technologies, n.d.). The ActivPAL recording must be initiated by the HCP via PAL Connect software prior to handing the device. For transtibial patients, the ActivPAL should be secured to the upper anterior thigh via adhesive bandages or PAL Stickies™ on the side of their affected leg. If the patient is bilateral, then either leg is a viable location. If the amputation level is transfemoral, the device should be secured to the prosthetic shaft in the same location as it would be on an intact leg. Some prosthetic legs may also have the ability to store the ActivPAL in the prosthetic shaft, as seen in Buis et al. (2014). The patient would then continuously wear the ActivPAL over the duration of

[1] Price estimated via amazon.co.uk, prices subject to change

a week. The ActivPAL can be given a plastic sock covering to prevent water damage, thus is suitable to be worn while showering, though it is not recommended to wear the device when submerged in water, for example during baths or swimming.

After a pre-determined amount of time (typically 7 days), the ActivPAL will automatically cease recording and switch itself off to conserve power. The HCP would then acquire the data from the patient when they return the device. The data (format: ".datx") is downloaded to a PC or laptop via PAL Connect software (PAL Technologies, Glasgow, UK) using a microUSB connector cable. They should then open the .datx file in PAL Analysis and export both the raw accelerometer data (uncompressed) and the Event data, which are both in .csv format.

To achieve the human activity recognition process of the framework, the author's proposed solution is to compile a downloadable .exe program formatted in Python using Tensorflow API. The program should intake the .csv files and output easily understandable graphical outputs regarding the patient's physical activity (see Section 9.1.5). Ideally, the HAR process should be directly integrated into PAL Analysis as a separate configurable output.

9.1.4 What is being measured?

If the clinical activity monitoring framework can provide a trained supervised classifier (currently a Support Vector Machine) generalizable to the ILLA population in the early stages of their prosthetic rehabilitation, then it would be able to recognize level walking, uphill and downhill movement, and upstairs and downstairs movement. With further improvement of the classifier, it could also potentially distinguish walking activities on hard and soft ground. Through ActivPAL's proprietary VANE algorithm, a measurement of step count (and subsequently cadence) and detection of sedentary and standing behaviour will also be provided. By extrapolating the average number of steps contained within a sampling window, it should also therefore be possible to specify level, hill, and stair movement into steps. If the HCP can provide input for their client's average stride length on level ground, slopes, and stairs, it should also be possible to provide an analysis of their walking speed.

If, however, a generalizable trained classifier cannot be provided, the clinical framework would still be able to recognize activities via an unsupervised classifier algorithm, but this would be relegated only to the detection of ground walking and stair movement, with currently no indication of the direction of stair movement. The additional clinical activity monitoring outputs via VANE do not require training and are therefore unaffected.

9.1.5 How is the data presented?

From the HCP's perspective, once data has been passed through the machine learning systems, the GUI should resemble the mock-ups depicted in Figs. 9.2-9.3. If the system were to be integrated into PAL Analysis, the colour aesthetic would also be altered to white and orange to have continuity with the main PAL software products. Due to the current inaccuracies of the supervised systems, a disclaimer should also be added stating the average accuracy of the system and including error bars in the step count graphs would be a worthwhile addition. If a trained supervised system was not practical to obtain due to a lack of resources, an unsupervised system would use a similar layout but only include level walking and stairs as measured outputs, perhaps complementing the step counts with the low dimensional tSNE models presented in Appendix E.

Once the output has been acquired by the HCP, they should deliver their feedback to the amputee in the method of their choosing. From Chapter 3.2.2.1, most HCPs appear to prefer to have feedback face-to-face, though by outputting the graphs as downloadable image files, the HCP can also elect to print or email the images to their clients with minimal effort. Following on from the discussion of the future of behavioural interventions in Chapter 2.4, it may be beneficial to incorporate modern technology into the framework by having the HCP contact their client via phone calls or text (with prior consent from the client) to follow-up on their rehabilitation progress. This could also help improve compliance with the ActivPAL by having the HCP ask their patient if they are wearing the device correctly.

Figure 9.2 Mock-up of GUI of physical activity data as presented to HCPs or ILLAs

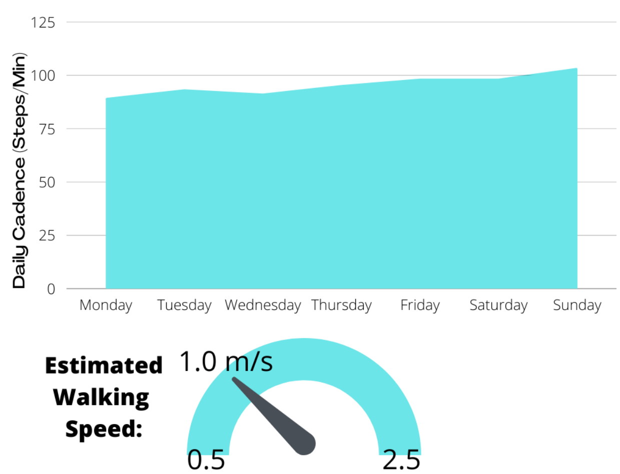


Figure 9.3 Additional GUI screens of the framework

9.1.6 Is the data reliable?

By discussing the reliability of the data, note that the word "reliability" is considered as a separate concept to "accuracy", which has been one of the main topics of the thesis. Here, reliability refers to whether the data acquired from a patient is trustworthy and cannot be manipulated via external means.

The ActivPAL can provide anti-tamper procedures to the HAR process by only allowing physical activity data to be analysed during periods of valid walking as recognized by the VANE algorithm. For example, shaking the ActivPAL vigorously will not be recognized as walking and therefore will not count towards data analysis. Unfortunately, there is no automated protection against tampering techniques such as simulating the movement of the ActivPAL in a walking motion while the patient holds it in their hand (which would be difficult to carry out), or by giving the ActivPAL to a friend or family member to wear instead. If the HCP suspects tampering, a potential solution could be to carry out a brief recording session with the patient under direct supervision and compare the accelerometer data collected under supervision to those gathered without supervision in PAL Analysis.

From a data-focused perspective, the ActivPAL is secure for its intended use. A PAL device can only be activated and deactivated by connecting the ActivPAL to a PC using a microUSB cable and using PAL Connect software, which also requires creating an account with PAL Technologies and downloading the appropriate software. This alone provides a major hurdle for the vast majority of the target population. Crucially, the core .datx file which contains the raw accelerometer readings is, while not fully encrypted, extremely difficult to decipher and manipulate in a meaningful way. For clinical trials however, the software should be updated to have some form of encryption.

9.1.7 How much does it cost?

Table 9.1 Capital expenditure for proposed clinical activity monitoring system

Capital Expenditure	Estimated Cost(£)	Source
ActivPAL 4	250.85	[Carpenter, 2021]†
PAL Software (PAL Analysis, PAL Connect)	0‡	[PAL Software]

† - Price additionally confirmed in email with PAL Technologies. Does not include tax and shipping.
‡ - Purchasing an ActivPAL device comes with a free lifetime access to PAL Software

Table 9.2 Operating expenditure for proposed clinical activity monitoring system

Operating Expenditure	Cost per annum (£)	Estimated quantity per 10 persons per annum	Source†
Isopropyl Alcohol	6	1x1L Bottle	Source Chemicals.com
Tegaderm™ Adhesive Patch	9	50x27cm^2 Patches	Amazon.co.uk

† - Prices accurate as of August 2021, subject to change

Tables 9.1-9.2 demonstrates the primary capital and operating expenditures required to utilise the activity monitoring framework. The majority of cost comes from acquiring an ActivPAL device, costing approximately £250 per device with an additional 2-year warranty for each device. While this is quadruple the price of the least expensive FitBit® model (see Section 9.1.2), the primary advantage of using the ActivPAL in the early stages of rehabilitation is the security of the data (see Section 9.1.6). The ActivPAL device also provides inherent security against theft over the FitBit® with its simplistic design: with the lack of external UI and unremarkable physical design, the majority of subjects will not view the device as valuable and are unlikely to steal or sell the device. This could also backfire; if the patient does not view the device as valuable, they may be more prone to misplacing or losing the device. Should theft or misplacement become a major issue, the ActivPAL hardware may need to updated with a GPS module to allow it to be tracked by the HCP. As Table 9.2 demonstrates, the operating expenditures needed for the framework are minimal, only requiring an adhesive to attach the ActivPAL in a secure fashion, and disinfectants when handing off or returning the ActivPAL device.

Finally, for any party who is interested in replicating the data collection methodology discussed in this thesis, Table 9.3 summarizes the primary costs required. These prices are accurate as of August 2021 and are subject to change depending on availability in the future. These costs are on top of the estimated capital and operating expenditures. The make of camera needed for data collection is unimportant; the only required properties are visible timestamps when reviewing footage and a battery life lasting at least one hour when video recording. It should also have external storage capacity (SD or MicroSD) to allow footage to be saved on other devices. To acquire elevation data, a refurbished iPhone 6 with a prepaid SIM card using a free Strava account is comparable or even less expensive than dedicated GPS loggers like a Garmin Tracker™. As stated in Chapter 5.2.2, using Strava with an iPhone 6 appears to be the most cost-effective method of obtaining accurate GPS data (Barberi, 2017). The disinfectant spray and screen wipes have been included presuming the intention to re-use the iPhone and camera strap with multiple participants.

Table 9.3 Estimated cost of replicating methodology

Data Collection costs	Price (£)	Source†
Camera	17	Amazon.co.uk
Camera Strap	10	Amazon.co.uk
32GB SD Card	6	Amazon.co.uk
Refurbished 16GB iPhone 6	79	giffgaff.com
100GB No-Contract SIM card	30	tescomobile.com
Dettol™ Cotton Disinfectant Spray 10 x 250ml	20	Amazon.co.uk
Screen Wipes	4	Amazon.co.uk
Total	**166**	-

† - Prices accurate as of August 2021, subject to change

9.1.8 Is the framework ethical?

As stated in the conclusion of Chapter 3.4, there will be many roadblocks which could prevent the implementation of the clinical activity monitoring system that are beyond the scope of this thesis to discuss (Chadwell et al., 2020). Nonetheless, if the participant is fully informed about the data collection process and agrees to the process both verbally and in writing, the "red-tape" associated with approving the clinical activity monitoring framework in a clinical setting should hopefully be minimal. The participant should be given clear and concise explanations about all stages of the data collection process, in particular, about who has access to the data. This should ideally be limited only to the HCP themselves. The data collection process does not require modification of the user's prosthesis, so as long as procedures are put in place to disinfect the ActivPAL device after use, there is no increased physical danger associated with wearing the ActivPAL.

9.1.9 Why is this important?

Perhaps the most vital component of the activity monitoring framework, and consequently the research presented in this thesis, is how it can be clinically relevant. In Chapter 1.3, the general importance of physical activity in an ILLA population was explained. Amputees who are physically active will see positive effects for their physical and mental wellbeing, including improvements in heart and lung functionality and self-esteem (Bragaru et al., 2011; Wetterhahn, Hanson, and Levy, 2002). Therefore, if a clinician can accurately track their patient's physical activity, they can acquire an objective understanding of the physical activity behaviour of their client and work towards improving their physical activity. In Chapter 2.7, it was stated that the most commonly used metric for physical activity measurement – step count – is too simplistic. A question that may arise from this statement is: what makes the activities investigated in this thesis different? After all, the primary measurement of the clinical activity monitoring system is still step count, but now broken down into more detail (uphill stepping, downstairs stepping etc.). The patient's rehabilitative progress is the key to this question: At the beginning of a patient's prosthetic rehabilitation, a patient may be reluctant to use their prosthesis, preferring to use more stable methods of transportation such as a wheelchair or crutches. As the patient becomes more comfortable using their prosthesis, this would be reflected in the monitoring system by the uptake of hill movement, stair movement, or in future iterations of the machine learning process, different types of terrains. The HCP can therefore utilize this system to track the progress of their clients in a much more meaningful way as opposed to using a simple pedometer or fitness tracker. These measurements can be combined with the standardized forms of physical function assessment described in Chapter 2.6 to provide a better understanding of the patient's rehabilitation progress. As many ILLA patients are elderly and/or have comorbidities which can impact on walking functionality, the framework does not concern itself with more vigorous activities such as running and jumping. Frankly, HCPs are unlikely to require quantitative monitoring of these activities, as ILLAs who are capable of such feats should have visibly apparent levels of physical fitness.

An important component not covered in great detail was regarding how the activity monitoring itself can lead to improvements in the patient's physical activity. This answer is deeply embedded in the field of psychology and to look into how HCP can use various tactics to encourage physical activity based on the data they are given is beyond the author's expertise and may in of itself be a separate thesis project. The primary intention of this clinical activity monitoring framework is to provide the HCP with an objective summation of their client's activity, from which they can make informed clinical decisions. For example, if a patient took ten-thousand steps on level ground, but very little or no steps up and down stairs, the HCP could use this clinical activity monitoring system to detect this issue, and from there discuss with the patient possible solutions as to how they could try and begin stair traversal, for example, by changing a component of the prosthetic, or through physiotherapy sessions.

9.2 Reflection of Work

9.2.1 Thesis Limitations and Future Work

The methodology applied in this thesis was satisfactory in terms of being to achieve its research goals, especially considering the restrictions imposed by the coronavirus epidemic. However, upon reflection there are a number of ways in which the methodology could have been improved. The limitations and suggested improvements for the methodology are addressed on a by-chapter basis beginning with Chapter 3. Chapter 1 is a simple introductory chapter in regard to the importance of physical activity in ILLA populations, and Chapter 2 is themed on physical activity interventions. While this second chapter was important for constructing the narrative towards building a clinical activity monitoring system, its focus differs greatly from the main theme of the thesis, which is based more around the HAR process.

A lack of participants from the interview stage greatly limited the generalizability of the thesis's findings, particularly in regard to desirable activity monitoring outcomes. More effort could have been put into the recruitment process. In Chapter 3.3.4, there was a brief mention of providing a reward incentive for participating, but an alternative less expensive route would have been to extend recruitment to beyond Glasgow and advertise recruitment across the whole of the UK, allowing participants to join in sessions via video conferencing in place of in-person appearances. There should also have been more targeted recruiting towards other stakeholders in a clinical activity monitoring system aside from prosthetists and physiotherapists, such as specialist research scientists and prosthetic manufacturers. The focus group sessions could have also been handled in a more focused and concise manner by contacting professional focus group companies and gathering advice and feedback on the proposed questions before conducting the interviews. The next stage of the research, at least in terms of acquiring professional opinions, should be to follow up on the HCPs initially interviewed, present the clinical activity framework, and consider their responses as a way of improving the framework. As the framework is aimed at collecting data from ILLAs in the early stages of their prosthetic rehabilitation, another focus group or even a simple interview or survey targeted at these demographics would provide invaluable feedback.

Due to a lack of machine learning experience prior to the conduction of the thesis, the methodological quality of the human activity recognition process was held back by the author's inexperience. Subsequently, much of the literature review of machine learning was spent investigating the many possible configurations of machine learning, including all the possible sensor modalities, feature extraction processes, dimensionality reduction processes as well as the many kinds of classifiers (supervised, unsupervised, and neural networks). While this process was necessary for the author to become familiar with the machine learning environment, it resulted in an admittedly shallow perspective of machine learning. The methodology of the thesis would have benefitted greatly if it had been known from the beginning of the thesis that it should be machine learning focused, thus giving time for the author to first familiarize themselves with machine learning, and by the time the data collection process was complete, they could have conceived intricately designed machine learning systems to try and acquire better classification rates. For an example of a more involved design, the author did not explore the possibility of using ensemble classifiers outside of Random Forests and AdaBoost. The weak learners in ensembles do not have to be limited to decision trees and can include “stronger” learners such as SVM and kNN which can give a better-informed classification decision in human activity recognition (Ni, Zhang, and Li, 2018).

Aside from the obvious lack of number and diversity of non-amputated and ILLA participants, which was largely unavoidable due to restrictions caused by the coronavirus epidemic, the data collection process would have benefitted from strongly encouraging the recruited participants to walk on more terrains to provide a more balanced dataset. Because of ethical restrictions, recruitment and data collection from a non-amputated population had to begin before collecting data from the ILLA population, as this provided evidence that the methodology carried out was safe to be replicated with a more vulnerable population. A more beneficial data collection process would have been to begin with data collection from the ILLA population, then instruct the non-amputated population to “match” the ILLA data collection by walking on the same types of terrains. While antithetical to the meaning of “free-living”, another action that should have been taken would be to have at least one instance of a non-amputated individual and an ILLA carry out the same walking route, this would eliminate any confounding factors relating to the traversed terrain and allow direct comparison of normal and prosthetic-aided activity data. In regard to the ambiguity of annotation, particularly with the identification of hill segments, future studies should require at least two independent annotators, ideally having an expert or experienced cohort in video annotation to ensure that all activity labels acquired are reliable. While it is not feasible to increase the accuracy of elevation data without the use of sophisticated equipment that would sharply increase the financial cost, a free solution to the identification of hill movement would be to ask participants to signal on the camera when they feel they are tackling an uphill or downhill slope: for example a thumbs-up could indicate uphill movement, a thumbs-down for downhill movement, and a handwave for when hill movement flattens out. While this process would not be entirely reliable in of itself due to relying on human perception, it would provide an additional source of input to the elevation data and would lead to more precise timestamp markings for the beginning and end of hill movement. Additional calibration procedures should be taken to ensure that the magnetometer signals are stabilized and subsequently can utilize the magnetometer features in machine learning.

Towards the end of Chapter 5, a lack of parameter tuning in regard to the pre-processing stage (e.g changing filter parameters, sample window size) was justified by the large amount of time required for feature calculation after each iteration of pre-processing. In future iterations of the algorithm, with ample reduction of the initial feature set, tuning of these parameters in conjunction with the classifier parameters should be possible via the implementation of Genetic Algorithm (GA) optimization. The GA, as its name suggests, is a process that seeks to optimize a system by performing random "mutations" (i.e alterations to tuning parameters) and selecting the best combinations of parameters, iterating the mutation and selection process until some pre-defined requirement is met (e.g maximum number of iterations, loss function reaching a threshold value). GA have been applied in HAR processes with some powerful classification results (Huang et al., 2021; Nguyen, Huynh, and Pham, 2018; Ozcan and Basturk, 2020).

In the supervised aspect of the HAR process, it was discovered that activity recognition was only feasible at Level 1 of resolution ("Level walking/flat", "uphill", "downhill", "upstairs" and "downstairs"). Ideally, future iterations of the algorithm should be able to achieve Level 2 of resolution with 8 walking activities ("Hard, flat", "Hard, uphill", "Hard, downhill", "Soft, flat", "Soft, uphill", "Soft, downhill", "upstairs" and "downstairs"). Some suggestions as to how this could be achieved include an improved version of SMOTE where there is further undersampling of the majority classes, a more refined technique for synthetic signal generation of raw data for the LSTM classifier, and the development of more intricately designed classifiers such as those discussed in Chapter 6.4.1 and this conclusive chapter. Features from the magnetometer should also help achieve this goal. More thought also needs to be put into the feature extraction process on a per-axis basis, instead of blindly including all features relating to each axis as well as features derived from the jerk and magnitude of the signal.

For the unsupervised aspect of the HAR process, the next crucial step is to validate the theory that appreciable differences in walking speed are the driving factor behind successful formation of the hill and stair clusters in low dimensional clustering models. The algorithm proposed in Chapter 7 will require further adaptation to differentiate upstairs and downstairs clusters, as well as handle Step-to ambulation methods and the different cluster formations generated by magnetometer features. UMAP should also be given further consideration as an alternative to the tSNE modelling process.

Another key change that needs to be made in the machine learning process is the integration of the project into a Python programming environment, using TensorFlow for machine learning processes, and Keras if using deep learning implementation. Throughout the supervised and unsupervised experimentations, it was found the Matlab environment was often insufficient in terms of the classification and tuning algorithms available. From observation, most state-of-the-art machine learning implementations are typically written in Python, and so would need to change the working environment to take advantage of these functions. Migration of the project is also mandatory for implementation within a clinical environment, as the cost for acquiring a Matlab license is far too expensive to be realistically implemented in a nation-wide setting. Python and TensorFlow, being open source, is free to use. For processing larger datasets, it would also be beneficial to take advantage of cloud processing. Databricks (Databricks inc., CA, USA) allows users to control computer clusters to process large quantities of data far faster than a standalone PC (Databricks, n.d.).

In terms of actions to be taken with the wearable device, there needs to be more exploration as to how measurements taken from ActivPAL's VANE analysis could acquire useful clinical activity monitoring outcomes. In particular, there will need to be validation of a patient's walking speed on level ground, slopes of different angles, and stairs. Validation of the wear time of the ActivPAL is also desirable and should be achievable via analysis of orientation data. While development of the VANE algorithm is the responsibility of PAL Technologies, some work can be undertaken to integrate the machine learning analysis into VANE to help acquire better accuracy of step counts during stair ambulation and the detection of sporadic stopping events. Vice versa, VANE analysis can also assist with the machine learning process by using the detection of standing or sitting events to filter out data that should not be recognized by a machine learning classifier.

There are some additional limitations of the research as a whole that does not particularly fit any single chapter: As is, the framework does not analyse physical activity data in realtime. Theoretically, it should be possible to acquire realtime monitoring first by installing a bluetooth module in the ActivPAL which would stream the data to a smartphone, which could then process and output activity predictions in realtime (Martin et al., 2017). A realtime system may help benefit the development of machine learning algorithms - for instance, researchers would be able to test various scenarios in laboratory or free-living conditions and have instant feedback as to whether an activity was correctly recognized, then use that feedback to help fine tune the algorithms. However, from the main demands of the stakeholders as discussed in Chapter 3, a realtime monitoring system has limited practical use in clincal settings. Many users of this framework would be elderly amputees who may not be technologically savvy, and so the number of potential ILLA users for this system is likely to be small. It could also be possible to stream the data to a server which would allow the HCPs to monitor their client's physical activity before they hand over the ActivPAL, but again this would have limited practical use; as identified in Chapter 3, HCPs are typically juggling their time between dozens of clients, and so likely would not have the time to be able to monitor their clients multiple times per week.

Another limitation of the research is the "transferability" of the framework that has been discussed in this thesis. The research presents a specific set of solutions to a specific problem, which raises the question as to whether changes in the paradigm to the problem or solution will still result in the same outcomes. A particularly questionable aspect of transferability is the wearable device: if the ActivPAL was replaced by for example a smartphone accelerometer, would the results presented in Chapters 6 & 7 change? Research conducted by Garcia-Gonzalez et al. (2020) demonstrates that even changes in sampling frequency from the same accelerometer device can impact on classification accuracies. If it is assumed that the smartphone contained in the accelerometer is exactly the same as the one in the ActivPAL (in terms of sampling frequency and ADC range), and is fixed to the participant in the same way that the ActivPAL was on the anterior thigh, results should theoretically be fairly similar. However, one must consider any smartphone will be larger and heavier than the ActivPAL, this could potentially influence the accelerometry behaviour during gait as the user would require slightly more energy to lift the phone. To the author's knowledge, no research has compared the ActivPAL to another accelerometer that is placed in the same location as the ActivPAL. Pfister et al. (2017) compared the ActivPAL to the ActiGraph™ GT3X+ to measure active and sedentary times and found weak agreement between the two devices, however it should be considered that the ActiGraph™ was worn on a waist belt while the ActivPAL was placed in the standard anterior thigh location, naturally producing different accelerometry signatures. While it would require validation, it seems unlikely that swapping the ActivPAL with another accelerometer will produce the same results without at minimum some alterations to the preprocessing methodology.

Continuing with the theme of "transferability", it is questionable whether the framework that was outlined in this thesis would be able to detect the same kinds of activities indoors - or distinguish stepping indoors from stepping outdoors. As stated in Chapter 5, it was not possible to monitor indoor activity in a laboratory environment due to the ongoing pandemic, and collecting accurate timestamps of activity in a participant's actual home without a camera to provide GTA would be problematic. As restrictions relating to the coronavirus have relaxed going into 2022 and beyond, it will be possible in future work to validate these activities indoors using laboratory environments. While research has demonstrated that transferring trained classifier systems from laboratory environments to free-living conditions will result in a drop of classification accuracy (Dutta et al., 2018; Pavey et al., 2017), there do not appear to be studies that have experimented with the reverse scenario. Intuitively, one would expect a similar outcome either way; in an indoor environment, the individual typically has more constrained movement in the mediolateral direction, and there would be more turning to navigate obstacles in the home. When it comes to stair traversal, public and domestic stairs have different regulations and requirements relating to step width and step height (Pear-Stairs, n.d.), which can result in different stepping behaviour indoors and outdoors. Private residences can construct spiral and alternating-tread staircases which require unique methods of traversal compared to an ordinary straight flight of stairs. These changes indicate that transferring the algorithms presented in this thesis to an indoor environment would be difficult, but still possible. Some additional training of systems on abberrant stair types (or distinguishing these types of steps as a different activity outcome altogether) may be required to allow full adapability of the system in indoor conditions.

The final component of transferability that merits discussion is to whether the system would work on different types of amputees not included in the study (for example, transfemoral amputees and amputees with total knee disarticulation) or individuals with other conditions that can impair movement, such as stroke or Parkinson's disease patients. As highlighted in Chapter 6, training classifiers on non-amputated patients did not transfer to recognition of activities carried out by ILLAs, and even by forcing the training model to only include ILLAs, there was a sharp drop in performance when the bilateral amputee was included in the training data. While the small number of participants makes it difficult to definitively state whether the system could transfer to other gait-impaired individuals, the results from this chapter strongly indicate that transfer learning would not be possible, and supervised classifiers must be built from the ground up with a particular group of gait-impaired individuals in mind. In regards to unsupervised learning however, the algorithm from Chapter 7 showed that it was possible to cluster and subsequently identify activity based only on data from the individual, therefore there is a distinct possibility that the algorithm used in that chapter could identify walking and stair data of stroke and Parkinson's disease patients.

A three-year hypothetical plan for further development of the activity monitoring system and framework is illustrated in Fig. 9.4. Immediately following completion of the thesis, a follow-up focus group would validate whether the activity monitoring system and the proposed framework from this thesis would benefit healthcare professionals, while some parallel research actions to be taken include migrating to an open-source Python environment, commence validation studies of activity monitoring outcomes that were identified in this thesis to be necessary to further improve the activity monitoring system. By six months, a brand new HAR study should be carried out, this time specifically targeting specific terrains to be traversed, as well as a more defined set of participants, with the aim of obtaining 10 non-amputated and 10 unilaterally amputated individuals - of which it would be further ideal to split between 5 transtibial and 5 transfemoral amputees. This would allow for further investigation of transfer learning studies as explored in experiment #4 of Chapter 6. After the follow-up human activity recognition study has been carried out and identification of activities from ILLAs can be reliably achieved, the next stage in development would be to once again evaluate the system with stakeholders to ensure they are satisfied with results. The aim would then be to implement the system into PAL's existing software, such as PAL analysis, followed by development of a standalone application or Application Programming Interfaces (APIs) that other software developers in the field of HAR could use to help enhance their research. The remaining years would then involve identifying and targeting relevant partners to the NHS who would benefit from this system and submit a tender for funding the development of ActivPAL devices (plus any other necessary hardware and software licensing), with the main goal being the commencement of the activity monitoring system in clinical trials and clinical practice after 2 years.

9.2.2 Overall Findings

The main objective of this thesis was to conduct novel exploration of physical activity data of non-amputated individuals and Individuals with lower limb amputation in free-living conditions for the purposes of outlining a clinical physical activity monitoring framework to be utilized by healthcare professionals specializing in the care of lower limb amputees. The thesis began with an introduction of the importance of physical activity for an ILLA population. A systematic literature review then looked at the applications of behavioural and prosthetic based interventions that have attempted to improve ILLA's physical activity. It was found that all identified studies when assessing intervention efficacy either used subjective measurements or objective measurements like step or activity counts, which was an oversimplification of the amputee's physical activity behaviour. At this point, the thesis pivoted away from researching the interventions themselves and focused more on the quality of the objective measurements used in the interventions. Thus began an effort to conceive a physical activity monitoring system that could capture clinically relevant physical activities in an ILLA population. This began with conducting a series of interviews and a focus group with the primary stakeholders of the system: HCPs (specialist physiotherapists and prosthetists) and experienced ILLAs. From these interviews, the desirable activity monitoring outcomes were determined as well as how the basic principles of the clinical activity monitoring framework should operate. In order to achieve useful activity monitoring measurements, it was necessary to utilize machine learning techniques, and so a systematic review of HAR was carried out. The principal findings of this review were that free-living studies of HAR

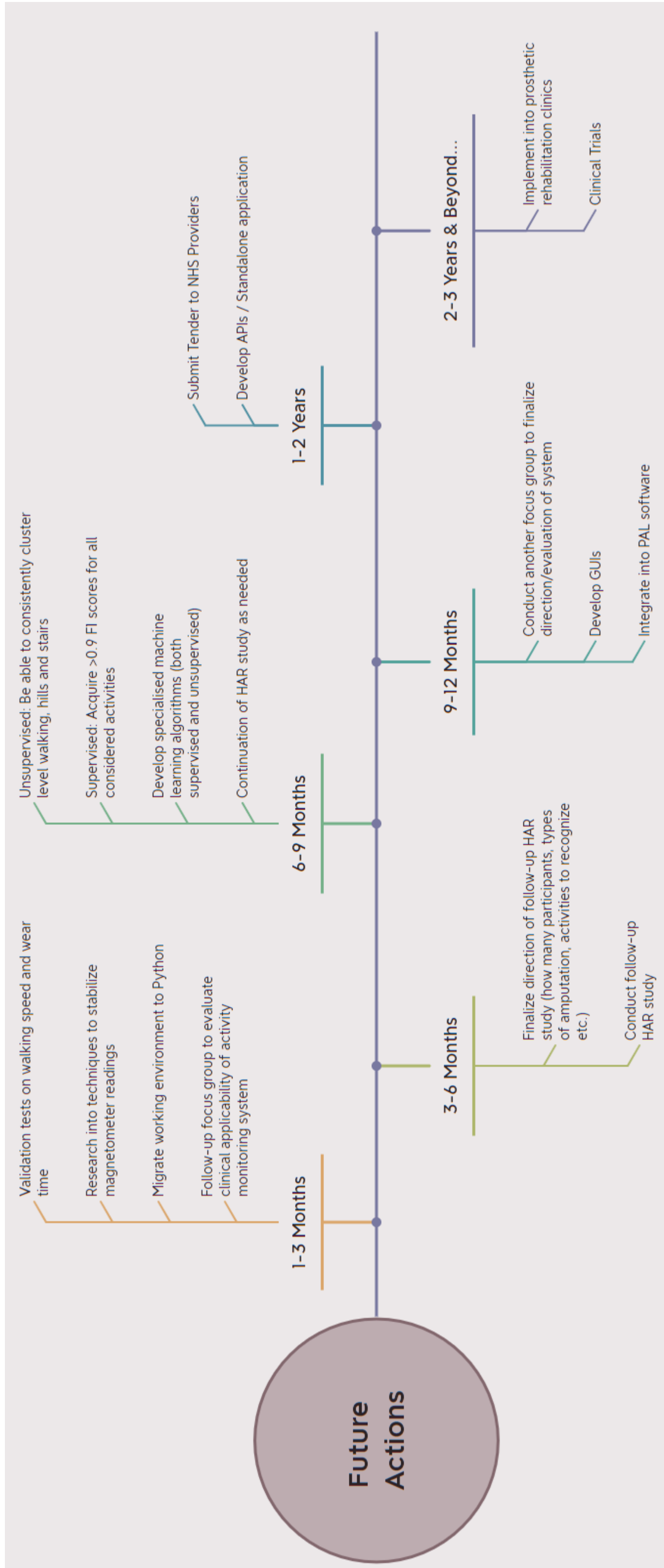


Figure 9.4 3 year timeline plan for further development of the clinical activity monitoring system

of an ILLA population are extremely rare, and no studies had attempted to carry the work out with an unsupervised machine learning approach. This co-incidentally benefitted the research direction due to the ongoing Covid-19 restrictions preventing any laboratory research from being conducted. Thus began the experimental stage of the thesis.

A group of ILLAs (n = 4) and non-amputated individuals (n = 8) were recruited to take part in a free-living study in which they would be expected to carry out walks in the vicinity of their home while wearing an ActivPAL device, additionally recording the walks with a chest-mounted camera and recording the elevation readings on Strava to provide a method of ground truth annotation. Once the data collection process was complete, the raw accelerometer data from the ActivPAL was annotated with its corresponding activity, removing any periods of non-movement from analysis. The remaining data was pre-processed with a bandpass filter and subsequently segmented into a series of 40 sample (2 second) windows, and an array of features from the time, frequency and wavelet domains were extracted per each segment. The subsequent analysis was then split into supervised and unsupervised investigations.

From the supervised approach, it was discovered that the mean recognition accuracy using five different activities (level walking, uphill, downhill, upstairs, downstairs) achievable was around 77% using either a Support Vector Machine or Long-Short Term Memory Deep Learning classifier. While the recognition accuracies were not outstanding, they were comparable with other studies that have been conducted in similar free-living settings. When trying to increase the "resolution" of the annotated label, there was a considerable recognition accuracy drop, thus from these results it was deemed best to only consider annotation at a basic level of detail. Attempts to train the classifiers on the non-amputated population and test on an ILLA population were also unfortunately unsuccessful, acquiring mean recognition accuracies of only 55%, indicating that the system could not be trained to recognize activity of an ILLA using only data from a non-amputated population. There was further indication that separate training models may be required for different kinds of lower limb amputation.

In the unsupervised approach, meaningful cluster models that clustered based on activity was only achievable when the data was separated by each individual, and only clustering of ground walking ("flat", "uphill" and "downhill" combined) and stairs ("upstairs" and "downstairs" were achievable. A novel algorithm based on posterior probabilities of a Gaussian Mixture Model and cluster compactness was devised to perform automatic recognition of hill and stair activities without the need for training or annotated data. This led to a statistically significant increase in Normalized Mutual Information compared to applying a standard DBSCAN algorithm. This indicates an unsupervised activity monitoring system is at least capable of recognizing stair ambulation.

Upon completion of the machine learning experimentation, a series of ActivPAL-focused experiments were also carried out. An analysis of the step count acquired by the ActivPAL's VANE algorithm revealed that it statistically significantly underreported step count compared to the ground truth, though had good interrater reliability. An analysis of standing time also highlighted a weakness in the VANE algorithm in being unable to detect brief spontaneous bouts of standing. Though not applicable for use in Chapters 6 and 7, a brief exploration of magnetometer-based features in the data revealed a significant increase in supervised recognition accuracy and an improvement in the quality of the clustering models. These could strongly benefit the clinical activity monitoring system in the future should a way of keeping magnetometer calibration consistent be devised. Using these findings, the thesis then concluded with an outline of a clinical activity monitoring framework, which would be ideally suited for tracking detailed step movement on level ground, hills, and stairs of an ILLA in the early stages of their prosthetic rehabilitation.

Bibliography

Abdallah, Amna, Rawan S Abdulsadig, and Magdi Amien (2019). “A comparative study on human loco-motor activity recognition using wearable sensors”. In: *2019 International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE)*. IEEE. DOI: 10.1109/iccceee46830.2019.9070840. URL: http://10.0.4.85/iccceee46830.2019.9070840%20https://dx.doi.org/10.1109/iccceee46830.2019.9070840.

Abedin, Alireza et al. (2020). “Towards Deep Clustering of Human Activities from Wearables”. In: *arXiv pre-print server*. URL: arxiv:2008.01659%20https://arxiv.org/abs/2008.01659.

Abo El-Maaty, Ayman M. and Amr G. Wassal (2018). “Hybrid GA-PCA Feature Selection Approach for Inertial Human Activity Recognition”. In: *2018 IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 1027–1032. DOI: 10.1109/SSCI.2018.8628702.

Agrawal, V. (2016). “Clinical Outcome Measures for Rehabilitation of Amputees – A Review”. In: *Phys Med Rehabil Int.* 3.2.

Ahmad, Naseer et al. (2014). “Lower limb amputation in England: prevalence, regional variation and relationship with revascularisation, deprivation and risk factors. A retrospective review of hospital data”. In: *Journal of the Royal Society of Medicine* 107.12, pp. 483–489. ISSN: 0141-0768. DOI: 10.1177/0141076814557301. URL: http://10.0.4.153/0141076814557301%20https://dx.doi.org/10.1177/0141076814557301.

Ahmed, Arif et al. (2017). “Prevalence of Phantom Limb Pain, Stump Pain, and Phantom Limb Sensation among the Amputated Cancer Patients in India: A Prospective, Observational Study”. In: *Indian journal of palliative care* 23.1, pp. 24–35. ISSN: 0973-1075 1998-3735. DOI: 10.4103/0973-1075.197944. URL: https://pubmed.ncbi.nlm.nih.gov/28216859%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5294433/.

Ainsworth, B. E. (2008). “How do I measure physical activity in my patients? Questionnaires and objective methods”. In: 43.1, pp. 6–9. ISSN: 0306-3674. DOI: 10.1136/bjsm.2008.052449. URL: https://dx.doi.org/10.1136/bjsm.2008.052449.

Ajanki, Antti (2007). *Example of k-nearest neighbour classification*. Accessed: 2021-09. URL: https://commons.wikimedia.org/w/index.php?curid=2170282.

Albert, Mark V. et al. (2013). “Monitoring Functional Capability of Individuals with Lower Limb Amputations Using Mobile Phones”. In: *PLoS ONE* 8.6, e65340. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0065340. URL: https://dx.doi.org/10.1371/journal.pone.0065340.

Alex (2020). *Feedforward Neural Networks and Multilayer Perceptrons*. Accessed: 2021-09. URL: https://boostedml.com/2020/04/feedforward-neural-networks-and-multilayer-perceptrons.html.

Ali, M. M. et al. (2013). “A contemporary comparative analysis of immediate postoperative prosthesis placement following below-knee amputation”. In: *Ann Vasc Surg* 27.8, pp. 1146–53. ISSN: 0890-5096. DOI: 10.1016/j.avsg.2012.10.031.

Allamy, Haider (2014). “Methods to avoid over-fitting and under-fitting in supervised machine learning (comparative study)”. In: *Computer Science, Communication & Instrumentation Devices*.

Aloulou, Hamdi et al. (2017). “Activity Recognition Enhancement Based on Ground-Truth: Introducing a New Method Including Accuracy and Granularity Metrics”. In: *Enhanced Quality of Life and Smart Living*. Springer International Publishing, pp. 87–98. DOI: 10.1007/978-3-319-66188-9_8. URL: http://10.0.3.239/978-3-319-66188-9%7B%5C_%7D8%20https://dx.doi.org/10.1007/978-3-319-66188-9%7B%5C_%7D8.

Altini, Marco (2015). *Dealing with imbalanced data: undersampling, oversampling and proper cross-validation*. Accessed: 2021-09. URL: https://www.marcoaltini.com/blog/dealing-with-imbalanced-data-undersampling-oversampling-and-proper-cross-validation.

Altun, Kerem and Billur Barshan (2010). “Human Activity Recognition Using Inertial/Magnetic Sensor Units”. In: Springer Berlin Heidelberg, pp. 38–51. DOI: 10.1007/978-3-642-14715-9_5. URL: https://dx.doi.org/10.1007/978-3-642-14715-9_5.

Amini, Amineh, Teh Wah, and Hadi Saboohi (2014). “On Density-Based Data Streams Clustering Algorithms: A Survey”. In: *Journal of Computer Science and Technology* 29, pp. 116–141. DOI: 10.1007/s11390-013-1416-3.

Aminian, Saeideh and Erica A Hinckson (2012). “Examining the validity of the ActivPAL monitor in measuring posture and ambulatory movement in children”. In: *International Journal of Behavioral Nutrition and Physical Activity* 9.1, p. 119. ISSN: 1479-5868. DOI: 10.1186/1479-5868-9-119. URL: http://10.0.4.162/1479-5868-9-119%20https://dx.doi.org/10.1186/1479-5868-9-119.

Andreu, Javier, Rashmi Dutta Baruah, and Plamen Angelov (2011). “Real time recognition of human activities from wearable sensors by evolving classifiers”. In: *2011 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2011)*. IEEE. DOI: 10.1109/fuzzy.2011.6007595. URL: http://10.0.4.85/fuzzy.2011.6007595%20https://dx.doi.org/10.1109/FUZZY.2011.6007595.

Anguita, Davide et al. (2012). “Human Activity Recognition on Smartphones Using a Multiclass Hardware-Friendly Support Vector Machine”. In: *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, pp. 216–223. DOI: 10.1007/978-3-642-35395-6_30. URL: http://10.0.3.239/978-3-642-35395-6%7B%5C_%7D30%20https://dx.doi.org/10.1007/978-3-642-35395-6%7B%5C_%7D30.

Anguita, Davide et al. (2013). “A public domain dataset for human activity recognition using smartphones”. In: *Esann*. Vol. 3, p. 3.

Al-Ani, Tarik, Quynh Trang Le Ba, and Eric Monacelli (2007). “On-line Automatic Detection of Human Activity in Home Using Wavelet and Hidden Markov Models Scilab Toolkits”. In: *2007 IEEE International Conference on Control Applications*. IEEE. DOI: 10.1109/cca.2007.4389278. URL: http://10.0.4.85/cca.2007.4389278%20https://dx.doi.org/10.1109/cca.2007.4389278.

Ankerst, Mihael et al. (1999). “OPTICS”. In: *ACM SIGMOD Record* 28.2, pp. 49–60. ISSN: 0163-5808. DOI: 10.1145/304181.304187. URL: http://10.0.4.121/304181.304187%20https://dx.doi.org/10.1145/304181.304187.

Aquilino, William S (1994). “Interview mode effects in surveys of drug and alcohol use: A field experiment”. In: *Public opinion quarterly* 58.2, pp. 210–240. ISSN: 1537-5331.

Arch (Schrank), Elisa et al. (2018). “Step count accuracy of StepWatch and FitBit One™ among individuals with a unilateral transtibial amputation”. In: *Prosthetics and Orthotics International* 42, p. 030936461876713. DOI: 10.1177/0309364618767138.

Ariza Colpas, Paola et al. (2020). “Unsupervised Human Activity Recognition Using the Clustering Approach: A Review”. In: *Sensors* 20.9, p. 2702. ISSN: 1424-8220. DOI: 10.3390/s20092702. URL: http://10.0.13.62/s20092702%20https://dx.doi.org/10.3390/s20092702.

Asbury, Jo-Ellen (1995). “Overview of Focus Group Research”. In: *Qualitative Health Research* 5.4, pp. 414–420. ISSN: 1049-7323. DOI: 10.1177/104973239500500402. URL: http://10.0.4.153/104973239500500402%20https://dx.doi.org/10.1177/104973239500500402.

Attal, Ferhat et al. (2015). *Physical Human Activity Recognition Using Wearable Sensors*. DOI: 10.3390/s151229858.

Awasthi, Samuya (2020). *SEVEN MOST POPULAR SVM KERNELS*. Accessed: 2021-09. URL: https://dataaspirant.com/svm-kernels/.

Balakrishnama, Suresh and Aravind Ganapathiraju (1998). “Linear discriminant analysis-a brief tutorial”. In: *Institute for Signal and information Processing* 18.1998, pp. 1–8.

Balk, E. et al. (2018). *Lower Limb Prostheses: Measurement Instruments, Comparison of Component Effects by Subgroups, and Long-Term Outcomes*. Agency for Healthcare Research and Quality, p. 2.

Balli, Serkan, Ensar Arif Sağbaş, and Musa Peker (2019). "Human activity recognition from smart watch sensor data using a hybrid of principal component analysis and random forest algorithm". In: *Measurement and Control* 52.1-2, pp. 37–45. ISSN: 0020-2940. DOI: 10.1177/0020294018813692. URL: http://10.0.4.153/0020294018813692%20https://dx.doi.org/10.1177/0020294018813692.

Banos, Oresti et al. (2014). "Window Size Impact in Human Activity Recognition". In: *Sensors* 14.4, pp. 6474–6499. ISSN: 1424-8220. DOI: 10.3390/s140406474. URL: http://10.0.13.62/s140406474%20https://dx.doi.org/10.3390/s140406474.

Bao, Ling and Stephen S Intille (2004). "Activity Recognition from User-Annotated Acceleration Data". In: ed. by Alois Ferscha and Friedemann Mattern. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 1–17. ISBN: 978-3-540-24646-6.

Barberi, Jeff (2017). *GPS Accuracy Test: GPS vs. Smartphone vs. Cyclocomputer (Round 2)*. Accessed: 2021-09. URL: https://www.singletracks.com/mtb-gear/gps-accuracy-gps-vs-smartphone-vs-cyclocomputer/.

Barmparas, Galinos et al. (2010). "Epidemiology of Post-Traumatic Limb Amputation: A National Trauma Databank Analysis". In: *The American surgeon* 76, pp. 1214–22.

Baroudi, Loubna et al. (2020). "Estimating Walking Speed in the Wild". In: *Frontiers in Sports and Active Living* 2. ISSN: 2624-9367. DOI: 10.3389/fspor.2020.583848. URL: http://10.0.13.61/fspor.2020.583848%20https://dx.doi.org/10.3389/fspor.2020.583848.

Batool, Mouazma, Ahmad Jalal, and Kibum Kim (2019). "Sensors Technologies for Human Activity Analysis Based on SVM Optimized by PSO Algorithm". In: *2019 International Conference on Applied and Engineering Mathematics (ICAEM)*. IEEE. DOI: 10.1109/icaem.2019.8853770. URL: http://10.0.4.85/icaem.2019.8853770%20https://dx.doi.org/10.1109/icaem.2019.8853770.

Baumgartner, R. (2011). "Knieexartikulation und transgenikuläre Amputation". In: *Operative Orthopädie und Traumatologie* 23.4, pp. 289–295. ISSN: 0934-6694. DOI: 10.1007/s00064-011-0041-y. URL: https://dx.doi.org/10.1007/s00064-011-0041-y.

Bayat, Akram, Marc Pomplun, and Duc A Tran (2014). "A Study on Human Activity Recognition Using Accelerometer Data from Smartphones". In: *Procedia Computer Science* 34, pp. 450–457. ISSN: 1877-0509. DOI: https://doi.org/10.1016/j.procs.2014.07.009. URL: http://www.sciencedirect.com/science/article/pii/S1877050914008643.

Beavers, Ian (2017). *The Case of the Misguided Gyro*. Accessed: 2022-02. URL: https://www.analog.com/en/analog-dialogue/raqs/raq-issue-139.html.

Beks, P. J. et al. (1995). "Peripheral arterial disease in relation to glycaemic level in an elderly Caucasian population: the Hoorn Study". In: *Diabetologia* 38.1, pp. 86–96. ISSN: 0012-186X. DOI: 10.1007/bf02369357. URL: https://dx.doi.org/10.1007/bf02369357.

Bellman, R (1966). “Dynamic Programming”. In: *Science* 153.3731, pp. 34–37. ISSN: 0036-8075. DOI: 10.1126/science.153.3731.34. URL: http://10.0.4.102/science.153.3731.34%20https://dx.doi.org/10.1126/science.153.3731.34.

Bengio, Y, P Simard, and P Frasconi (1994). “Learning long-term dependencies with gradient descent is difficult”. In: *IEEE Transactions on Neural Networks* 5.2, pp. 157–166. ISSN: 1045-9227. DOI: 10.1109/72.279181. URL: http://10.0.4.85/72.279181%20https://dx.doi.org/10.1109/72.279181.

Bengio, Yoshua (2012). “Practical recommendations for gradient-based training of deep architectures”. In: *arXiv pre-print server*. DOI: None. URL: arxiv:1206.5533%20https://arxiv.org/abs/1206.5533.

Beravs, Tadej et al. (2014). “Magnetometer Calibration Using Kalman Filter Covariance Matrix for Online Estimation of Magnetic Field Orientation”. In: *IEEE Transactions on Instrumentation and Measurement* 63.8, pp. 2013–2020. ISSN: 0018-9456. DOI: 10.1109/tim.2014.2302240. URL: https://dx.doi.org/10.1109/tim.2014.2302240.

Bergmann, J H M and A H McGregor (2011). “Body-Worn Sensor Design: What Do Patients and Clinicians Want?” In: *Annals of Biomedical Engineering* 39.9, pp. 2299–2312. ISSN: 1573-9686. DOI: 10.1007/s10439-011-0339-9. URL: https://doi.org/10.1007/s10439-011-0339-9.

Bergstra, James and Yoshua Bengio (2012). “Random Search for Hyper-Parameter Optimization”. In: *Journal of Machine Learning Research* 13.10, pp. 281–305. URL: http://jmlr.org/papers/v13/bergstra12a.html.

Berman, Elena et al. (2019). “Maximizing precision and accuracy of the doubly labeled water method via optimal sampling protocol, calculation choices, and incorporation of 17O measurements”. In: *European Journal of Clinical Nutrition*. DOI: 10.1038/s41430-019-0492-z.

Biagetti, Giorgio et al. (2018). “Human activity monitoring system based on wearable sEMG and accelerometer wireless sensor nodes”. In: *BioMedical Engineering OnLine* 17.S1. ISSN: 1475-925X. DOI: 10.1186/s12938-018-0567-4. URL: https://dx.doi.org/10.1186/s12938-018-0567-4.

Blagus, Rok and Lara Lusa (2013). “SMOTE for high-dimensional class-imbalanced data”. In: *BMC Bioinformatics* 14.1, p. 106. ISSN: 1471-2105. DOI: 10.1186/1471-2105-14-106. URL: http://10.0.4.162/1471-2105-14-106%20https://dx.doi.org/10.1186/1471-2105-14-106.

Boone, David and Kim Coleman (2006). “Use of the Prosthesis Evaluation Questionnaire (PEQ)”. In: *JPO: Journal of Prosthetics and Orthotics* 18, P68–P79. DOI: 10.1097/00008526-200601001-00008.

Bosch (n.d.). *Pressure sensor BMP280*. Accessed: 2021-09. URL: https://www.bosch-sensortec.com/products/environmental-sensors/pressure-sensors/bmp280/.

Bouchard, Claude et al. (1983). "A method to assess energy expentiture in children and adults". In: *Am J Clin Nutr* 37, pp. 461–467. DOI: 10.1093/ajcn/37.3.461. URL: https://dx.doi.org/10.1093/ajcn/37.3.461.

Bourke, Alan, Espen Ihlen, and Jorunn Helbostad (2019). "Validation of the activPAL in Free-Living and Laboratory Scenarios for the Measurement of Physical Activity, Stepping, and Transitions in Older Adults". In: pp. 1–8. DOI: 10.1123/jmpb.2018-0056.

Bragaru, Mihai et al. (2013). "Barriers and Facilitators of Participation in Sports: A Qualitative Study on Dutch Individuals with Lower Limb Amputation". In: *PLoS ONE* 8.3, e59881. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0059881. URL: https://dx.doi.org/10.1371/journal.pone.0059881.

Bragaru, Mihail et al. (2011). "Amputees and Sports". In: *Sports Medicine* 41.9, pp. 721–740. ISSN: 1179-2035. DOI: 10.2165/11590420-000000000-00000. URL: https://doi.org/10.2165/11590420-000000000-00000.

Bravata, Dena M. et al. (2007). "Using Pedometers to Increase Physical Activity and Improve Health". In: *JAMA* 298.19, p. 2296. ISSN: 0098-7484. DOI: 10.1001/jama.298.19.2296.

Bro, Rasmus and Age K Smilde (2014). "Principal component analysis". In: *Anal. Methods* 6.9, pp. 2812–2831. ISSN: 1759-9660. DOI: 10.1039/c3ay41907j. URL: http://10.0.4.15/c3ay41907j%20https://dx.doi.org/10.1039/c3ay41907j.

Brooks, Dina et al. (2001). "The 2-minute walk test as a measure of functional improvement in persons with lower limb amputation". In: *Archives of Physical Medicine and Rehabilitation* 82.10, pp. 1478–1483. ISSN: 0003-9993. DOI: 10.1053/apmr.2001.25153. URL: https://dx.doi.org/10.1053/apmr.2001.25153.

Bubnis, Daniel (2018). *How to Calculate Stride Length and Step Length*. Accessed: 2021-09. URL: https://www.healthline.com/health/stride-length#calculate-step-and-stride-length.

Buis, Arjan et al. (2014). "Measuring the Daily Stepping Activity of People with Transtibial Amputation Using the ActivPAL™ Activity Monitor". In: *JPO Journal of Prosthetics and Orthotics* Volume 26, pp. 43–47. DOI: 10.1097/JPO.0000000000000016.

Bujang, Mohamad Adam and N Baharum (2017). "Guidelines of the minimum sample size requirements for Cohen's Kappa". In: *Epidemiology Biostatistics and Public Health* 14, e12267–1. DOI: 10.2427/12267.

Al-Busaidi, Zakiya Q (2008). "Qualitative research and its uses in health care". In: *Sultan Qaboos University medical journal* 8.1, pp. 11–19. ISSN: 2075-051X. URL: https:

//pubmed.ncbi.nlm.nih.gov/21654952%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3087733/.

Bussmann, Johannes B., Eleonore A. Grootscholten, and Henk J. Stam (2004). "Daily physical activity and heart rate response in people with a unilateral transtibial amputation for vascular disease". In: *Archives of Physical Medicine and Rehabilitation* 85.2, pp. 240–244. ISSN: 0003-9993. DOI: 10.1016/s0003-9993(03)00485-4. URL: https://dx.doi.org/10.1016/s0003-9993(03)00485-4.

Byun, Seonjeong et al. (2019). "Walking-speed estimation using a single inertial measurement unit for the older adults". In: *PLOS ONE* 14.12, e0227075. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0227075. URL: http://10.0.5.91/journal.pone.0227075%20https://dx.doi.org/10.1371/journal.pone.0227075.

Carlson, Jordan A et al. (2021). "Validity of Two Awake Wear-Time Classification Algorithms for activPAL in Youth, Adults, and Older Adults". In: *Journal for the Measurement of Physical Behaviour* 4.2, pp. 151–162. DOI: 10.1123/jmpb.2020-0045. URL: https://journals.humankinetics.com/view/journals/jmpb/4/2/article-p151.xml.

Carpenter, Chelsea, Chih-Hsiang Yang, and Delia West (2021). *A Comparison of Sedentary Behavior as Measured by the Fitbit and ActivPAL in College Students*. DOI: 10.3390/ijerph18083914.

Castro, Oscar et al. (2018). "A scoping review on interventions to promote physical activity among adults with disabilities". In: *Disability and Health Journal* 11.2, pp. 174–183. ISSN: 1936-6574. DOI: https://doi.org/10.1016/j.dhjo.2017.10.013. URL: http://www.sciencedirect.com/science/article/pii/S193665741730211X.

Chadwell, Alix et al. (2020). "Technology for monitoring everyday prosthesis use: a systematic review". In: *Journal of NeuroEngineering and Rehabilitation* 17.1. ISSN: 1743-0003. DOI: 10.1186/s12984-020-00711-4. URL: http://10.0.4.162/s12984-020-00711-4%20https://dx.doi.org/10.1186/s12984-020-00711-4.

Chambers, G S et al. (2002). "Hierarchical recognition of intentional human gestures for sports video annotation". In: *Object recognition supported by user interaction for service robots*. IEEE Comput. Soc. DOI: 10.1109/icpr.2002.1048493. URL: http://10.0.4.85/icpr.2002.1048493%20https://dx.doi.org/10.1109/icpr.2002.1048493.

Chaurasia, S and S R N Reddy (2018). "Design and Implementation of Data Collection & Analysis Tool for Healthcare Parameter Monitoring using Inverse Low Pass Filter". In: *EAI Endorsed Transactions on Pervasive Health and Technology* 4.16, p. 160460. ISSN: 2411-7145. DOI: 10.4108/eai.30-10-2018.160460. URL: http://10.0.16.12/eai.30-10-2018.160460%20https://dx.doi.org/10.4108/eai.30-10-2018.160460.

Chawla, Nitesh et al. (2002). "SMOTE: Synthetic Minority Over-sampling Technique". In: *J. Artif. Intell. Res. (JAIR)* 16, pp. 321–357. DOI: 10.1613/jair.953.

Chen, Kong Y. et al. (2012). “Redefining the roles of sensors in objective physical activity monitoring”. In: *Medicine and science in sports and exercise* 44.1 Suppl 1, S13–S23. ISSN: 1530-0315 0195-9131. DOI: 10.1249/MSS.0b013e3182399bc8. URL: https://pubmed.ncbi.nlm.nih.gov/22157770%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3245644/.

Chen, Yuqing and Yang Xue (2015). “A Deep Learning Approach to Human Activity Recognition Based on Single Accelerometer”. In: *2015 IEEE International Conference on Systems, Man, and Cybernetics*. IEEE. DOI: 10.1109/smc.2015.263. URL: http://10.0.4.85/smc.2015.263%20https://dx.doi.org/10.1109/SMC.2015.263.

Chen, Yuwen et al. (2016). “LSTM Networks for Mobile Human Activity Recognition”. In: *Proceedings of the 2016 International Conference on Artificial Intelligence: Technologies and Applications*. Atlantis Press. DOI: 10.2991/icaita-16.2016.13. URL: http://10.0.11.175/icaita-16.2016.13%20https://dx.doi.org/10.2991/icaita-16.2016.13.

Chen, Zhenghua et al. (2017). “Robust Human Activity Recognition Using Smartphone Sensors via CT-PCA and Online SVM”. In: *IEEE Transactions on Industrial Informatics* 13.6, pp. 3070–3080. ISSN: 1551-3203. DOI: 10.1109/tii.2017.2712746. URL: http://10.0.4.85/tii.2017.2712746%20https://dx.doi.org/10.1109/TII.2017.2712746.

Chen, Zhenghua et al. (2018). “Distilling the Knowledge From Handcrafted Features for Human Activity Recognition”. In: *IEEE Transactions on Industrial Informatics* 14.10, pp. 4334–4342. ISSN: 1551-3203. DOI: 10.1109/tii.2018.2789925. URL: http://10.0.4.85/tii.2018.2789925%20https://dx.doi.org/10.1109/TII.2018.2789925.

Chi, Eric C and Kenneth Lange (2015). “Splitting Methods for Convex Clustering”. In: *Journal of Computational and Graphical Statistics* 24.4, pp. 994–1013. ISSN: 1061-8600. DOI: 10.1080/10618600.2014.948181. URL: http://10.0.4.56/10618600.2014.948181%20https://dx.doi.org/10.1080/10618600.2014.948181.

Chin, Takaaki et al. (2002). “Physical Fitness of Lower Limb Amputees”. In: *American Journal of Physical Medicine & Rehabilitation* 81.5, pp. 321–325. DOI: 10.1097/00002060-200205000-00001.

Christiansen, Cory L. et al. (2015). “Functional Outcomes After the Prosthetic Training Phase of Rehabilitation After Dysvascular Lower Extremity Amputation”. In: *PM & R : the journal of injury, function, and rehabilitation* 7.11, pp. 1118–1126. ISSN: 1934-1563 1934-1482. DOI: 10.1016/j.pmrj.2015.05.006. URL: https://pubmed.ncbi.nlm.nih.gov/25978948%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4643436/.

Christiansen, Cory L. et al. (2018). “Behavior-Change Intervention Targeting Physical Function, Walking, and Disability After Dysvascular Amputation: A Randomized Controlled Pilot Trial”. In: *Archives of Physical Medicine and Rehabilitation* 99.11, pp. 2160–2167. ISSN: 0003-9993. DOI: 10.1016/j.apmr.2018.04.011. URL: https://dx.doi.org/10.1016/j.apmr.2018.04.011.

Cindy Ng, Li Whye, Sue Jenkins, and Kylie Hill (2012). “Accuracy and responsiveness of the stepwatch activity monitor and ActivPAL in patients with COPD when walking with and without a rollator”. In: *Disability and Rehabilitation* 34.15, pp. 1317–1322. ISSN: 0963-8288. DOI: 10.3109/09638288.2011.641666. URL: http://10.0.12.37/09638288.2011.641666%20https://dx.doi.org/10.3109/09638288.2011.641666.

Clark, M. (2018). “Compensatory hip and knee mechanics in transtibial amputees during stair descent and directional task”. MA thesis. Northern Michigan University. URL: https://commons.nmu.edu/theses/547.

Cohen, Jacob (1960). “A Coefficient of Agreement for Nominal Scales”. In: *Educational and Psychological Measurement* 20.1, pp. 37–46. ISSN: 0013-1644. DOI: 10.1177/001316446002000104. URL: http://10.0.4.153/001316446002000104%20https://dx.doi.org/10.1177/001316446002000104.

Cola, Guglielmo, Alessio Vecchio, and Marco Avvenuti (2014). “Improving the performance of fall detection systems through walk recognition”. In: 5.6, pp. 843–855. ISSN: 1868-5137. DOI: 10.1007/s12652-014-0235-x. URL: http://10.0.3.239/s12652-014-0235-x%20https://dx.doi.org/10.1007/s12652-014-0235-x.

Coleman, K. L. et al. (2004). “Quantification of prosthetic outcomes: elastomeric gel liner with locking pin suspension versus polyethylene foam liner with neoprene sleeve suspension”. In: *J Rehabil Res Dev* 41.4, pp. 591–602. ISSN: 0748-7711. DOI: 10.1682/jrrd.2004.04.0591.

Cross, Rod (1999). “Standing, walking, running, and jumping on a force plate”. In: *American Journal of Physics* 67.4, pp. 304–309. ISSN: 0002-9505. DOI: 10.1119/1.19253. URL: http://10.0.4.95/1.19253%20https://dx.doi.org/10.1119/1.19253.

Cruciani, Federico et al. (2019). “A Public Domain Dataset for Human Activity Recognition in Free-Living Conditions”. In: *2019 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*. IEEE. DOI: 10.1109/smartworld-uic-atc-scalcom-iop-sci.2019.00071. URL: http://10.0.4.85/smartworld-uic-atc-scalcom-iop-sci.2019.00071%20https://dx.doi.org/10.1109/SmartWorld-UIC-ATC-SCALCOM-IOP-SCI.2019.00071.

Daines, Kyle J. F. et al. (2021). “Fall risk classification for people with lower extremity amputations using random forests and smartphone sensor features from a 6-minute walk test”. In: *PLOS ONE* 16.4. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0247574.

Dale, R Barry (2012). “Clinical Gait Assessment”. In: *Physical Rehabilitation of the Injured Athlete*. Elsevier, pp. 464–479. DOI: 10.1016/b978-1-4377-2411-0.00021-6. URL: http://10.0.3.248/b978-1-4377-2411-0.00021-6%20https://dx.doi.org/10.1016/b978-1-4377-2411-0.00021-6.

Databricks (n.d.). Accessed: 2022-02. URL: https://docs.databricks.com/clusters/configure.html#cluster-size-and-autoscaling.

Davie-Smith, F., J. Hebenton, and H. Scott (2018). "A Survey of the Lower Limb Amputee Population in Scotland 2015 Full Report". In: *Scottish Physiotherapy Amputee Research Group*.

Davies, S. C. et al. (2019). "UK Chief Medical Officers' Physical Activity Guidelines". In: *Department of Health and Social Care*.

Day, Melissa Catherine, Ross Wadey, and Siobhan Strike (2019). "Living with limb loss: everyday experiences of "good" and "bad" days in people with lower limb amputation". In: *Disability and Rehabilitation* 41.20, pp. 2433–2442. ISSN: 0963-8288. DOI: 10.1080/09638288.2018.1467502. URL: http://10.0.4.56/09638288.2018.1467502%20https://dx.doi.org/10.1080/09638288.2018.1467502.

De, Debraj et al. (2015). "Multimodal Wearable Sensing for Fine-Grained Activity Recognition in Healthcare". In: *IEEE Internet Computing* 19.5, pp. 26–35. ISSN: 1089-7801. DOI: 10.1109/mic.2015.72. URL: http://10.0.4.85/mic.2015.72%20https://dx.doi.org/10.1109/MIC.2015.72.

Deans, Sarah et al. (2012). "Motivations and barriers to prosthesis users participation in physical activity, exercise and sport: A review of the literature". In: *Prosthetics and orthotics international* 36, pp. 260–9. DOI: 10.1177/0309364612437905.

Deans, Sarah et al. (2020). "Reliability and Criterion-Related Validity of the activPAL? Accelerometer When Measuring Physical Activity and Sedentary Behavior in Adults With Lower Limb Absence". In: *Journal for the Measurement of Physical Behaviour* 3.3, pp. 244–252. DOI: 10.1123/jmpb.2019-0045. URL: https://journals.humankinetics.com/view/journals/jmpb/3/3/article-p244.xml.

Deathe, Barry, William C. Miller, and Mark Speechley (2002). "The status of outcome measurement in amputee rehabilitation in Canada". In: *Archives of Physical Medicine and Rehabilitation* 83.7, pp. 912–918. ISSN: 0003-9993. DOI: https://doi.org/10.1053/apmr.2002.33221. URL: http://www.sciencedirect.com/science/article/pii/S0003999302000059.

Dehghani, Akbar et al. (2019). "A Quantitative Comparison of Overlapping and Non-Overlapping Sliding Windows for Human Activity Recognition Using Inertial Sensors". In: *Sensors* 19.22, p. 5026. ISSN: 1424-8220. DOI: 10.3390/s19225026. URL: http://10.0.13.62/s19225026%20https://dx.doi.org/10.3390/s19225026.

Dehzangi, Omid and Vaishali Sahu (2018). "IMU-Based Robust Human Activity Recognition using Feature Analysis, Extraction, and Reduction". In: *2018 24th International Conference on Pattern Recognition (ICPR)*, pp. 1402–1407. DOI: 10.1109/ICPR.2018.8546311.

Dekker, Rienk et al. (2018). "Pre-operative rehabilitation for dysvascular lower-limb amputee patients: A focus group study involving medical professionals". In: *PloS one* 13.10, e0204726–e0204726. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0204726. URL: https://pubmed.ncbi.nlm.nih.gov/30321178%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6188752/.

Delehanty, Rosalyn D. and Luxie Trachsel (1995). "Effects of short-term group treatment on rehabilitation outcome of adults with amputations". In: *International Journal of Rehabilitation and Health* 1.2, pp. 61–73. ISSN: 1068-9591. DOI: 10.1007/bf02213887. URL: https://dx.doi.org/10.1007/BF02213887.

Demrozi, Florenc et al. (2020). "Human Activity Recognition Using Inertial, Physiological and Environmental Sensors: A Comprehensive Survey". In: *IEEE Access* 8, pp. 210816–210836. ISSN: 2169-3536. DOI: 10.1109/access.2020.3037715. URL: http://10.0.4.85/access.2020.3037715%20https://dx.doi.org/10.1109/access.2020.3037715.

Dengel, Andreas et al. (2016). *Human Activity Recognition: Using Sensor Data of Smartphones and Smartwatches*. DOI: 10.5220/0005816004880493.

Despois, Julien (2018). *What is the difference between a convolutional neural network and a multilayer perceptron?* Accessed: 2021-09. URL: https://www.quora.com/What-is-the-difference-between-a-convolutional-neural-network-and-a-multilayer-perceptron.

Devlin, M. et al. (2004). "Houghton Scale of prosthetic use in people with lower-extremity amputations: Reliability, validity, and responsiveness to change". In: *Arch Phys Med Rehabil* 85.8, pp. 1339–44. ISSN: 0003-9993 (Print) 0003-9993. DOI: 10.1016/j.apmr.2003.09.025.

Dhammi, Ish Kumar and Sudhir Kumar (2014). "Osteosarcoma: A journey from amputation to limb salvage". In: *Indian journal of orthopaedics* 48.3, pp. 233–234. ISSN: 0019-5413 1998-3727. DOI: 10.4103/0019-5413.132486. URL: https://pubmed.ncbi.nlm.nih.gov/24932025%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4052018/.

Dillingham, T. R., L. E. Pezzin, and E. J. MacKenzie (2002). "Limb amputation and limb deficiency: epidemiology and recent trends in the United States". In: *South Med J* 95.8, pp. 875–83. ISSN: 0038-4348 (Print) 0038-4348. DOI: 10.1097/00007611-200208000-00018.

Ding, C and H Peng (2003). "Minimum redundancy feature selection from microarray gene expression data". In: *Computational Systems Bioinformatics. CSB2003. Proceedings of the 2003 IEEE Bioinformatics Conference. CSB2003*. IEEE Comput. Soc. DOI: 10.1109/csb.2003.1227396. URL: http://10.0.4.85/csb.2003.1227396%20https://dx.doi.org/10.1109/csb.2003.1227396.

Dinis, J (2018). "Hierarchical Classification using hierarchical clustering: an application to Human Activity Recognition". PhD thesis.

DoJ, ed. (2010). *2010 ADA standards for accessible design*.

Dowd, Kieran P, Deirdre M Harrington, and Alan E Donnelly (2012). "Criterion and Concurrent Validity of the activPAL™ Professional Physical Activity Monitor in Adolescent Females". In: *PLoS ONE* 7.10, e47633. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0047633. URL: http://10.0.5.91/journal.pone.0047633%20https://dx.doi.org/10.1371/journal.pone.0047633.

Dutta, Arindam et al. (2018). "Identifying Free-Living Physical Activities Using Lab-Based Models with Wearable Accelerometers". In: *Sensors* 18.11, p. 3893. ISSN: 1424-8220. DOI: 10.3390/s18113893.

Edwardson, Charlotte L et al. (2017). "Considerations when using the activPAL monitor in field-based research with adult populations". In: *Journal of Sport and Health Science* 6.2, pp. 162–178. ISSN: 2095-2546. DOI: 10.1016/j.jshs.2016.02.002. URL: http://10.0.3.248/j.jshs.2016.02.002%20https://dx.doi.org/10.1016/j.jshs.2016.02.002.

Efron, B (1979). "Bootstrap Methods: Another Look at the Jackknife". In: *The Annals of Statistics* 7.1, pp. 1–26. ISSN: 0090-5364. DOI: 10.1214/aos/1176344552. URL: http://10.0.4.190/aos/1176344552%20https://dx.doi.org/10.1214/aos/1176344552.

El Moudden, Ismail et al. (2018). *Learned Model For Human Activity Recognition Based On Dimensionality Reduction*.

Ellis, Katherine et al. (2014). "Multi-sensor physical activity recognition in free-living". In: *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*. ACM. DOI: 10.1145/2638728.2641673. URL: http://10.0.4.121/2638728.2641673%20https://dx.doi.org/10.1145/2638728.2641673.

Engenheiro, Gonçalo et al. (Oct. 2020). "Falls in Unilateral Lower Limb Amputees Living in the Community: A Portuguese Study". In: *Acta medica portuguesa* 33. DOI: 10.20344/amp.12615.

English, Coralie et al. (2016). "Sitting and Activity Time in People With Stroke". In: *Physical Therapy* 96.2, pp. 193–201. ISSN: 0031-9023. DOI: 10.2522/ptj.20140522. URL: https://doi.org/10.2522/ptj.20140522.

Erdaş, Ç.Berke et al. (2016). "Integrating Features for Accelerometer-based Activity Recognition". In: *Procedia Computer Science* 98, pp. 522–527. ISSN: 1877-0509. DOI: https://doi.org/10.1016/j.procs.2016.09.070. URL: https://www.sciencedirect.com/science/article/pii/S1877050916322153.

Ermes, M et al. (2008). "Detection of Daily Activities and Sports With Wearable Sensors in Controlled and Uncontrolled Conditions". In: *IEEE Transactions on Information Technology in Biomedicine* 12.1, pp. 20–26. ISSN: 1089-7771. DOI: 10.1109/titb.2007.899496. URL: http://10.0.4.85/titb.2007.899496%20https://dx.doi.org/10.1109/TITB.2007.899496.

Espinilla, Macarena et al. (2018). *Human Activity Recognition from the Acceleration Data of a Wearable Device. Which Features Are More Relevant by Activities?* DOI: 10.3390/proceedings2191242.

Esquenazi, A. (2004). "Amputation rehabilitation and prosthetic restoration. From surgery to community reintegration". In: *Disabil Rehabil* 26.14-15, pp. 831–6. ISSN: 0963-8288 (Print) 0963-8288. DOI: 10.1080/09638280410001708850.

Fay, Michael P and Michael A Proschan (2010). "Wilcoxon-Mann-Whitney or t-test? On assumptions for hypothesis tests and multiple interpretations of decision rules". In: *Statistics Surveys* 4.none, pp. 1–39. ISSN: 1935-7516. DOI: 10.1214/09-ss051. URL: http://10.0.4.190/09-ss051%20https://dx.doi.org/10.1214/09-ss051.

Fearn, Tom (2010). "Double Cross-Validation". In: *NIR news* 21.5, pp. 14–15. ISSN: 0960-3360. DOI: 10.1255/nirn.1194. URL: http://10.0.4.231/nirn.1194%20https://dx.doi.org/10.1255/nirn.1194.

Figo, Davide et al. (2010). "Preprocessing techniques for context recognition from accelerometer data". In: *Personal and Ubiquitous Computing* 14.7, pp. 645–662. ISSN: 1617-4909. DOI: 10.1007/s00779-010-0293-9. URL: http://10.0.3.239/s00779-010-0293-9%20https://dx.doi.org/10.1007/s00779-010-0293-9.

Finch, W. (2019). "A Comparison of Clustering Methods when Group Sizes are Unequal, Outliers are Present, and in the Presence of Noise Variables". In: *General Linear Model Journal* 45.1, pp. 12–22.

Fitbit (n.d.[a]). Accessed: 2022-02. URL: https://help.fitbit.com/articles/en_US/Help_article/1141.htm#:~:text=Fitbit%5C%20devices%5C%20that%5C%20count%5C%20floors,take%5C%20to%5C%20calculate%5C%20floors%5C%20climbed..

— (n.d.[b]). Accessed: 2021-09. URL: https://www.fitbit.com/global/uk/products.

Flores, Andrew Christian et al. (2018). "An Evaluation of SVM and Naive Bayes with SMOTE on Sentiment Analysis Data Set". In: *2018 International Conference on Engineering, Applied Sciences, and Technology (ICEAST)*. IEEE. DOI: 10.1109/iceast.2018.8434401. URL: http://10.0.4.85/iceast.2018.8434401%20https://dx.doi.org/10.1109/iceast.2018.8434401.

Fogelberg, Donald J et al. (2016). "What People Want in a Prosthetic Foot: A Focus Group Study". In: *JPO Journal of Prosthetics and Orthotics* 28.4, pp. 145–151. ISSN: 1040-8800. DOI: 10.1097/jpo.0000000000000102. URL: http://10.0.4.73/jpo.0000000000000102%20https://dx.doi.org/10.1097/jpo.0000000000000102.

Franchignoni, Franco et al. (2004). "Reliability, validity, and responsiveness of the locomotor capabilities index in adults with lower-limb amputation undergoing prosthetic training11No commercial party having a direct financial interest in the results of the research supporting this articl". In: *Archives of Physical Medicine and Rehabilitation*

85.5, pp. 743–748. ISSN: 0003-9993. DOI: 10.1016/j.apmr.2003.06.010. URL: https://dx.doi.org/10.1016/j.apmr.2003.06.010.

Franti, Pasi and Sami Sieranoja (2019). "How much can k-means be improved by using better initialization and repeats?" In: *Pattern Recognition* 93, pp. 95–112. ISSN: 0031-3203. DOI: https://doi.org/10.1016/j.patcog.2019.04.014. URL: https://www.sciencedirect.com/science/article/pii/S0031320319301608.

Frick, Eric Christopher (2015). "Mitigation of magnetic interference and compensation of bias drift in inertial sensors". In: DOI: 10.17077/etd.cetctkqj. URL: https://dx.doi.org/10.17077/etd.cetctkqj.

Frølich, Laura and Irene Dowding (2018). "Removal of muscular artifacts in EEG signals: a comparison of linear decomposition methods". In: *Brain Informatics* 5.1, pp. 13–22. ISSN: 2198-4018. DOI: 10.1007/s40708-017-0074-6. URL: http://10.0.3.239/s40708-017-0074-6%20https://dx.doi.org/10.1007/s40708-017-0074-6.

Fujiyama, Taku and Nick Tyler (2004). "An explicit study on walking speeds of pedestrians on stairs". In: *10th International Conference on Mobility and Transport for Elderly and Disabled People*.

Fukuoka, Yoshimi et al. (2015). "A Novel Diabetes Prevention Intervention Using a Mobile App". In: 49.2, pp. 223–237. ISSN: 0749-3797. DOI: 10.1016/j.amepre.2015.01.003. URL: https://dx.doi.org/10.1016/j.amepre.2015.01.003.

Fullerton, Elliott, Ben Heller, and Mario Munoz-Organero (2017). "Recognizing Human Activity in Free-Living Using Multiple Body-Worn Accelerometers". In: *IEEE Sensors Journal* 17.16, pp. 5290–5297. ISSN: 1530-437X. DOI: 10.1109/jsen.2017.2722105. URL: http://10.0.4.85/jsen.2017.2722105%20https://dx.doi.org/10.1109/jsen.2017.2722105.

Furber, Susan et al. (2008). "The effectiveness of a brief intervention using a pedometer and step-recording diary in promoting physical activity in people diagnosed with type 2 diabetes or impaired glucose tolerance". In: *Health Promotion Journal of Australia* 19.3, pp. 189–195. ISSN: 1036-1073. DOI: 10.1071/he08189.

Fushiki, Tadayoshi (2011). "Estimation of prediction error by using K-fold cross-validation". In: *Statistics and Computing* 21.2, pp. 137–146. ISSN: 0960-3174. DOI: 10.1007/s11222-009-9153-8. URL: http://10.0.3.239/s11222-009-9153-8%20https://dx.doi.org/10.1007/s11222-009-9153-8.

Gailey, R. S. et al. (2002). "The amputee mobility predictor: an instrument to assess determinants of the lower-limb amputee's ability to ambulate". In: *Arch Phys Med Rehabil* 83.5, pp. 613–27. ISSN: 0003-9993 (Print) 0003-9993. DOI: 10.1053/ampr.2002.32309.

Gallagher, Pamela and Malcolm Maclachlan (2004). "The Trinity amputation and prosthesis experience scales and quality of life in people with lower-limb amputation". In: 85.5, pp. 730–736. ISSN: 0003-9993. DOI: 10.1016/j.apmr.2003.07.009. URL: https://dx.doi.org/10.1016/j.apmr.2003.07.009.

Gani, Md Osman et al. (2019). “A light weight smartphone based human activity recognition system with high accuracy”. In: *Journal of Network and Computer Applications* 141, pp. 59–72. ISSN: 1084-8045. DOI: 10.1016/j.jnca.2019.05.001. URL: http://10.0.3.248/j.jnca.2019.05.001%20https://dx.doi.org/10.1016/j.jnca.2019.05.001.

Gao, Lei, A K Bourke, and John Nelson (2014). “Evaluation of accelerometer based multi-sensor versus single-sensor activity recognition systems”. In: *Medical Engineering & Physics* 36.6, pp. 779–785. ISSN: 1350-4533. DOI: https://doi.org/10.1016/j.medengphy.2014.02.012. URL: https://www.sciencedirect.com/science/article/pii/S1350453314000344.

Garber, Carol Ewing et al. (2010). “Physical and mental health-related correlates of physical function in community dwelling older adults: a cross sectional study”. In: *BMC Geriatrics* 10.1, p. 6. ISSN: 1471-2318. DOI: 10.1186/1471-2318-10-6. URL: https://dx.doi.org/10.1186/1471-2318-10-6.

Garcia-Ceja, Enrique et al. (2014). “Long-Term Activity Recognition from Wristwatch Accelerometer Data”. In: 14.12, pp. 22500–22524. ISSN: 1424-8220. DOI: 10.3390/s141222500. URL: http://10.0.13.62/s141222500%20https://dx.doi.org/10.3390/s141222500.

Garcia-Gonzalez, Daniel et al. (2020). “A Public Domain Dataset for Real-Life Human Activity Recognition Using Smartphone Sensors”. In: *Sensors (Basel, Switzerland)* 20.8, p. 2200. ISSN: 1424-8220. DOI: 10.3390/s20082200. URL: https://pubmed.ncbi.nlm.nih.gov/32295028%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7218897/.

Gardner, David et al. (2016). “Monitoring Prosthesis User Activity and Doffing Using an Activity Monitor and Proximity Sensors”. In: *Journal of Prosthetics and Orthotics* 28, pp. 68–77. DOI: 10.1097/JPO.0000000000000093.

Gardner, S.J. et al. (2011). *Amputation as a Last Resort*. URL: https://evtoday.com/articles/2011-aug/amputation-as-a-last-resort.

Gates, Deanna H et al. (2012). “Gait characteristics of individuals with transtibial amputations walking on a destabilizing rock surface”. In: *Gait & Posture* 36.1, pp. 33–39. ISSN: 0966-6362. DOI: 10.1016/j.gaitpost.2011.12.019. URL: http://10.0.3.248/j.gaitpost.2011.12.019%20https://dx.doi.org/10.1016/j.gaitpost.2011.12.019.

Ghazali, Mohd Fazli (2018). “Awareness, potential factors, and post-amputation care of stump flexion contractures among transtibial amputees”. In: *Turkish Journal of Physical Medicine and Rehabilitation* 64.3, pp. 268–276. ISSN: 2587-0823. DOI: 10.5606/tftrd.2018.1668. URL: https://dx.doi.org/10.5606/tftrd.2018.1668.

Gjoreski, Hristijan and Matjaz Gams (2011). “Accelerometer data preparation for activity recognition”. In: *Artificial Intelligence and Ambient Intelligence*. URL: https://www.researchgate.net/publication/259340203.

Golestani, Negar and Mahta Moghaddam (2020). “Human activity recognition using magnetic induction-based motion signals and deep recurrent neural networks”. In: *Nature Communications* 11.1. ISSN: 2041-1723. DOI: 10.1038/s41467-020-15086-2. URL: https://dx.doi.org/10.1038/s41467-020-15086-2.

Google-Developers (2020). *Clustering Algorithms*. Accessed: 2021-09. URL: https://developers.google.com/machine-learning/clustering/clustering-algorithms.

Gordon, M. et al. (2019). “App Usage Predicts Cognitive Ability in Older Adults”. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - paper 168* (Glasgow, United Kingdom). ACM. DOI: 10.1145/3290605.3300398. URL: https://doi.org/10.1145/3290605.3300398.

Görür, Dilan and Carl Edward Rasmussen (2010). “Dirichlet Process Gaussian Mixture Models: Choice of the Base Distribution”. In: *Journal of Computer Science and Technology* 25.4, pp. 653–664. ISSN: 1000-9000. DOI: 10.1007/s11390-010-9355-8. URL: http://10.0.3.239/s11390-010-9355-8%20https://dx.doi.org/10.1007/s11390-010-9355-8.

Grant, P. M. et al. (2006). “The validation of a novel activity monitor in the measurement of posture and motion during everyday activities”. In: *British Journal of Sports Medicine* 40.12, pp. 992–997. ISSN: 0306-3674. DOI: 10.1136/bjsm.2006.030262. URL: https://dx.doi.org/10.1136/bjsm.2006.030262.

Grayson, Mark, Kevin Shatzkamer, and Scott Wainner (2009). *IP design for mobile networks*. Indianapolis, IN: Cisco Press, pp. 354–362. ISBN: 9781587058264.

Grbich, Carol (1999). *Qualitative research in health: An introduction.* Thousand Oaks, CA: Sage Publications Ltd, p. 312. ISBN: 0-7619-6103-8.

Guo, Junqi et al. (2016). “Smartphone-Based Patients’ Activity Recognition by Using a Self-Learning Scheme for Medical Monitoring.” In: *Journal of medical systems* 40.6, p. 140. ISSN: 1573-689X (Electronic). DOI: 10.1007/s10916-016-0497-2.

Guo, Xinjian et al. (2008). “On the Class Imbalance Problem”. In: *2008 Fourth International Conference on Natural Computation*. IEEE. DOI: 10.1109/icnc.2008.871. URL: http://10.0.4.85/icnc.2008.871%20https://dx.doi.org/10.1109/icnc.2008.871.

Gupta, Piyush and Tim Dallas (2014). “Feature Selection and Activity Recognition System Using a Single Triaxial Accelerometer”. In: *IEEE Transactions on Biomedical Engineering* 61.6, pp. 1780–1786. ISSN: 0018-9294. DOI: 10.1109/tbme.2014.2307069. URL: http://10.0.4.85/tbme.2014.2307069%20https://dx.doi.org/10.1109/TBME.2014.2307069.

Gupta, Prankit, Richard Mcclatchey, and Praminda Caleb-Solly (2020). “Tracking changes in user activity from unlabelled smart home sensor data using unsupervised learning methods”. In: *Neural Computing and Applications* 32.16, pp. 12351–12362. ISSN: 0941-0643. DOI: 10.1007/s00521-020-04737-6. URL: http://10.0.3.239/s00521-020-04737-6%20https://dx.doi.org/10.1007/s00521-020-04737-6.

Gusfield, Dan (1997). *Algorithms on Strings, Trees, and Sequences: Computer Science and Computational Biology*. Cambridge: Cambridge University Press. ISBN: 9780521585194. DOI: DOI:10.1017/CBO9780511574931. URL: https://www.cambridge.org/core/books/algorithms-on-strings-trees-and-sequences/F0B095049C7E6EF5356F0A26686C20D3.

Gyllensten, I C and A G Bonomi (2011). "Identifying Types of Physical Activity With a Single Accelerometer: Evaluating Laboratory-trained Algorithms in Daily Life". In: *IEEE Transactions on Biomedical Engineering* 58.9, pp. 2656–2663. ISSN: 0018-9294. DOI: 10.1109/tbme.2011.2160723. URL: http://10.0.4.85/tbme.2011.2160723%20https://dx.doi.org/10.1109/tbme.2011.2160723.

Hafner, Brian J. et al. (2007). "Evaluation of Function, Performance, and Preference as Transfemoral Amputees Transition From Mechanical to Microprocessor Control of the Prosthetic Knee". In: *Archives of Physical Medicine and Rehabilitation* 88.2, pp. 207–217. ISSN: 0003-9993. DOI: 10.1016/j.apmr.2006.10.030. URL: https://dx.doi.org/10.1016/j.apmr.2006.10.030.

Hamer, Mark et al. (2020). "Feasibility of Measuring Sedentary Time Using Data From a Thigh-Worn Accelerometer". In: *American Journal of Epidemiology* 189.9, pp. 963–971. ISSN: 0002-9262. DOI: 10.1093/aje/kwaa047. URL: http://10.0.4.69/aje/kwaa047%20https://dx.doi.org/10.1093/aje/kwaa047.

Han, Jin Tae et al. (2009). "Three-Dimensional Kinematic Analysis during Upslope Walking with Different Inclinations by Healthy Adults". In: *Journal of Physical Therapy Science* 21.4, pp. 385–391. ISSN: 0915-5287. DOI: 10.1589/jpts.21.385. URL: http://10.0.6.53/jpts.21.385%20https://dx.doi.org/10.1589/jpts.21.385.

Hanley, Marisol A. et al. (2007). "Preamputation Pain and Acute Pain Predict Chronic Pain After Lower Extremity Amputation". In: *The Journal of Pain* 8.2, pp. 102–109. ISSN: 1526-5900. DOI: https://doi.org/10.1016/j.jpain.2006.06.004. URL: http://www.sciencedirect.com/science/article/pii/S152659000600887X.

Harding, Jessica L. et al. (2020). "National and State-Level Trends in Nontraumatic Lower-Extremity Amputation Among U.S. Medicare Beneficiaries With Diabetes, 2000–2017". In: *Diabetes Care* 43.10, pp. 2453–2459. ISSN: 0149-5992. DOI: 10.2337/dc20-0586.

Harper, Nicole G., Jason M. Wilken, and Richard R. Neptune (2017). "Muscle Function and Coordination of Stair Ascent". In: *Journal of Biomechanical Engineering* 140.1. ISSN: 0148-0731. DOI: 10.1115/1.4037791. URL: https://doi.org/10.1115/1.4037791.

Hashmi, Muhammad Zeeshan Ul Hasnain et al. (2019). "What Lies Beneath One's Feet? Terrain Classification Using Inertial Data of Human Walk". In: *Applied Sciences* 9.15, p. 3099. ISSN: 2076-3417. DOI: 10.3390/app9153099. URL: http://10.0.13.62/app9153099%20https://dx.doi.org/10.3390/app9153099.

Hassan, Mohammed Mehedi et al. (2018). "A robust human activity recognition system using smartphone sensors and deep learning". In: *Future Generation Computer Systems*

81, pp. 307–313. ISSN: 0167-739X. DOI: 10.1016/j.future.2017.11.029. URL: http://10.0.3.248/j.future.2017.11.029%20https://dx.doi.org/10.1016/j.future.2017.11.029.

Hawkins, Alexander T. et al. (2016). "The effect of social integration on outcomes after major lower extremity amputation". In: *Journal of vascular surgery* 63.1, pp. 154–162. ISSN: 1097-6809 0741-5214. DOI: 10.1016/j.jvs.2015.07.100. URL: https://pubmed.ncbi.nlm.nih.gov/26474508%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4739523/.

Heary, Caroline and Eilis Hennessy (2012). "Focus Groups Versus Individual Interviews with Children: A Comparison of Data". In: *The Irish Journal of Psychology* 27, pp. 58–68. DOI: 10.1080/03033910.2006.10446228.

Hebert, J. S. et al. (2009). "Outcome measures in amputation rehabilitation: ICF body functions". In: *Disabil Rehabil* 31.19, pp. 1541–54. ISSN: 0963-8288 (Print) 0963-8288. DOI: 10.1080/09638280802639467.

Hergenroeder, Andrea L et al. (2018). "Accuracy of Objective Physical Activity Monitors in Measuring Steps in Older Adults". In: *Gerontology and Geriatric Medicine* 4, p. 233372141878112. ISSN: 2333-7214. DOI: 10.1177/2333721418781126. URL: http://10.0.4.153/2333721418781126%20https://dx.doi.org/10.1177/2333721418781126.

Highsmith, M Jason et al. (2010). "Differences in the Spatiotemporal Parameters of Transtibial and Transfemoral Amputee Gait". In: *JPO: Journal of Prosthetics and Orthotics* 22.1. ISSN: 1040-8800. URL: https://journals.lww.com/jpojournal/Fulltext/2010/01000/Differences%7B%5C_%7Din%7B%5C_%7Dthe%7B%5C_%7DSpatiotemporal%7B%5C_%7DParameters%7B%5C_%7Dof.5.aspx.

Hinton, Geoffrey and Sam Roweis (2003). "Stochastic Neighbor Embedding". In: *Advances in Neural Information Processing Systems*. Vol. 15.

Hintz, Kasper S, Henrik Vedel, and Eigil Kaas (2019). "Collecting and processing of barometric data from smartphones for potential use in numerical weather prediction data assimilation". In: *Meteorological Applications* 26.4, pp. 733–746. ISSN: 1350-4827. DOI: 10.1002/met.1805. URL: http://10.0.3.234/met.1805%20https://dx.doi.org/10.1002/met.1805.

Hobara, Hiroaki et al. (2011). "Lower extremity joint kinematics of stair ascent in transfemoral amputees". In: *Prosthetics & Orthotics International* 35.4, pp. 467–472. ISSN: 0309-3646. DOI: 10.1177/0309364611425564. URL: http://10.0.4.153/0309364611425564%20https://dx.doi.org/10.1177/0309364611425564.

Hobs (n.d.). *How to choose the number of hidden layers and nodes in a feedforward neural network?* Accessed: 2021-09. URL: https://stats.stackexchange.com/q/136542.

Hochreiter, Sepp and Jürgen Schmidhuber (1997). "Long Short-Term Memory". In: *Neural Comput.* 9.8, pp. 1735–1780. ISSN: 0899-7667. DOI: 10.1162/neco.1997.9.8.1735. URL: https://doi.org/10.1162/neco.1997.9.8.1735.

Hof, A L, J P Van Zandwijk, and M F Bobbert (2002). "Mechanics of human triceps surae muscle in walking, running and jumping". In: *Acta Physiologica Scandinavica* 174.1, pp. 17–30. ISSN: 0001-6772. DOI: 10.1046/j.1365-201x.2002.00917.x. URL: http://10.0.4.22/j.1365-201x.2002.00917.x%20https://dx.doi.org/10.1046/j.1365-201x.2002.00917.x.

Hong, Yu-Jin et al. (2010). "Mobile health monitoring system based on activity recognition using accelerometer". In: *Simulation Modelling Practice and Theory* 18.4, pp. 446–455. ISSN: 1569-190X. DOI: 10.1016/j.simpat.2009.09.002. URL: http://10.0.3.248/j.simpat.2009.09.002%20https://dx.doi.org/10.1016/j.simpat.2009.09.002.

Hood, Sarah et al. (2020). "A kinematic and kinetic dataset of 18 above-knee amputees walking at various speeds". In: *Scientific Data* 7.

Hopper, Robert (1992). *Telephone conversation.* Vol. 724. Indiana University Press. ISBN: 025320724X.

Hordacre, Brenton, Christopher Barr, and Maria Crotty (2015). "Community activity and participation are reduced in transtibial amputee fallers: a wearable technology study". In: *BMJ Innovations* 1.1, p. 10. DOI: 10.1136/bmjinnov-2014-000014. URL: http://innovations.bmj.com/content/1/1/10.abstract.

Htike, Kyaw Kyaw et al. (2014). "Human activity recognition for video surveillance using sequences of postures". In: *The Third International Conference on e-Technologies and Networks for Development (ICeND2014)*, pp. 79–82. DOI: 10.1109/ICeND.2014.6991357.

Hu, B et al. (2018). "Machine learning algorithms based on signals from a single wearable inertial sensor can detect surface- and age-related differences in walking". In: *Journal of Biomechanics* 71, pp. 37–42. ISSN: 0021-9290. DOI: 10.1016/j.jbiomech.2018.01.005. URL: http://10.0.3.248/j.jbiomech.2018.01.005%20https://dx.doi.org/10.1016/j.jbiomech.2018.01.005.

Huan, Zhan et al. (2019). "Gait Recognition of Acceleration Sensor for Smart Phone Based on Multiple Classifier Fusion". In: *Mathematical Problems in Engineering* 2019, pp. 1–17. ISSN: 1024-123X. DOI: 10.1155/2019/6471532. URL: http://10.0.4.131/2019/6471532%20https://dx.doi.org/10.1155/2019/6471532.

Huang, Yu-Chuan et al. (2017). "A study on multiple wearable sensors for activity recognition". In: *2017 IEEE Conference on Dependable and Secure Computing*, pp. 449–452. DOI: 10.1109/DESEC.2017.8073827.

Huang, Guangbin et al. (2012). "Extreme Learning Machine for Regression and Multiclass Classification". In: *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 42, pp. 513–529.

Huang, Zhenzhen et al. (2021). "Acceleration Feature Extraction of Human Body Based on Wearable Devices". In: *Energies* 14.4, p. 924. ISSN: 1996-1073. DOI: 10.3390/en14040924. URL: http://10.0.13.62/en14040924%20https://dx.doi.org/10.3390/en14040924.

Husted, Hannah and Tamra Llewellyn (Jan. 2017). "The Accuracy of Pedometers in Measuring Walking Steps on a Treadmill in College Students". In: *International Journal of Exercise Science* 10, pp. 146–153.

Huynh, Tam (2008). "Human Activity Recognition with Wearable Sensors". PhD thesis.

Ignatov, Andrey D and Vadim V Strijov (2016). "Human activity recognition using quasiperiodic time series collected from a single tri-axial accelerometer". In: *Multimedia Tools and Applications* 75.12, pp. 7257–7270. ISSN: 1380-7501. DOI: 10.1007/s11042-015-2643-0. URL: http://10.0.3.239/s11042-015-2643-0%20https://dx.doi.org/10.1007/s11042-015-2643-0.

Inoue, Masaya, Sozo Inoue, and Takeshi Nishida (2018). "Deep recurrent neural network for mobile human activity recognition with high throughput". In: *Artificial Life and Robotics* 23.2, pp. 173–185. ISSN: 1433-5298. DOI: 10.1007/s10015-017-0422-x. URL: http://10.0.3.239/s10015-017-0422-x%20https://dx.doi.org/10.1007/s10015-017-0422-x.

UC-Irvine (n.d.). Accessed: 2022-01. URL: https://archive.ics.uci.edu/ml/index.php.

Ivan, Morgun (2018). *Density-based clustering in R*. Accessed: 2021-09. URL: https://en.proft.me/2017/02/3/density-based-clustering-r/.

Jamieson, Alexander (2021). *Dataset for: "Dataset for: Construction of a Clinical Activity Monitoring Framework Based on Free-living Investigations of Individuals with Lower Limb Amputation"*. DOI: 10.15129/ae451315-5258-4a07-8eb4-204e4d2e357f.

Jamieson, Alexander, Laura Murray, and Arjan Buis (2020). "The use of physical activity outcomes in rehabilitation interventions for lower limb amputees: a systematic review". In: *Canadian Prosthetics & Orthotics Journal* 3.1. ISSN: 2561-987X. DOI: 10.33137/cpoj.v3i1.33931. URL: https://dx.doi.org/10.33137/cpoj.v3i1.33931.

Jayakaran, Prasath, Meredith Perry, and Leigh Hale (2019). "Comparison of self-reported physical activity levels and quality of life between individuals with dysvascular and non-dysvascular below-knee amputation: A cross-sectional study". In: *Disability and Health Journal* 12.2, pp. 235–241. ISSN: 1936-6574. DOI: 10.1016/j.dhjo.2018.10.005. URL: https://dx.doi.org/10.1016/j.dhjo.2018.10.005.

Jiang, Wenchao and Zhaozheng Yin (2015). "Human Activity Recognition Using Wearable Sensors by Deep Convolutional Neural Networks". In: *Proceedings of the 23rd ACM International Conference on Multimedia*. New York, NY, USA: Association for Computing Machinery, pp. 1307–1310. ISBN: 9781450334594. DOI: 10.1145/2733373.2806333. URL: https://doi.org/10.1145/2733373.2806333.

Joshua, Liju and Koshy Varghese (2011). “Accelerometer-Based Activity Recognition in Construction”. In: *Journal of Computing in Civil Engineering* 25.5, pp. 370–379. ISSN: 0887-3801. DOI: 10.1061/(asce)cp.1943-5487.0000097. URL: http://10.0.4.37/(asce)cp.1943-5487.0000097%20https://dx.doi.org/10.1061/(asce)cp.1943-5487.0000097.

Jude, E. B. et al. (2001). “Peripheral Arterial Disease in Diabetic and Nondiabetic Patients: A comparison of severity and outcome”. In: *Diabetes Care* 24.8, pp. 1433–1437. ISSN: 0149-5992. DOI: 10.2337/diacare.24.8.1433. URL: https://dx.doi.org/10.2337/diacare.24.8.1433.

Jung, Im (2020). “A review of privacy-preserving human and human activity recognition”. In: *International Journal on Smart Sensing and Intelligent Systems* 13, pp. 1–13. DOI: 10.21307/ijssis-2020-008.

Kafle, Sabin and Dejing Dou (2016). “A Heterogeneous Clustering Approach for Human Activity Recognition”. In: *Big Data Analytics and Knowledge Discovery*. Springer International Publishing, pp. 68–81. DOI: 10.1007/978-3-319-43946-4_5. URL: http://10.0.3.239/978-3-319-43946-4%7B%5C_%7D5%20https://dx.doi.org/10.1007/978-3-319-43946-4%7B%5C_%7D5.

Kaghyan, Sahak and H. Sarukhanyan (2012). “Activity Recognition using K-nearest neighbor algorithm on smartphone with tri-axial accelerometer”. In: *International Journal Information Models and Analyses* 1.2, pp. 146–156.

Kanjo, Eiman, Eman M G Younis, and Chee Siang Ang (2019). “Deep learning analysis of mobile physiological, environmental and location sensor data for emotion detection”. In: *Information Fusion* 49, pp. 46–56. ISSN: 1566-2535. DOI: 10.1016/j.inffus.2018.09.001. URL: http://10.0.3.248/j.inffus.2018.09.001%20https://dx.doi.org/10.1016/j.inffus.2018.09.001.

Karagiannaki, Katerina, Athanasia Panousopoulou, and Panagiotis Tsakalides (2017). “An online feature selection architecture for Human Activity Recognition”. In: *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. DOI: 10.1109/icassp.2017.7952611. URL: http://10.0.4.85/icassp.2017.7952611%20https://dx.doi.org/10.1109/ICASSP.2017.7952611.

Karamizadeh, Sasan et al. (2013). “An Overview of Principal Component Analysis”. In: *Journal of Signal and Information Processing* 04.03, pp. 173–175. ISSN: 2159-4465. DOI: 10.4236/jsip.2013.43b031.

Kästner, Marika, Marc Strickert, and Thomas Villmann (2013). *A sparse kernelized matrix learning vector quantization model for human activity recognition. BT - 21st European Symposium on Artificial Neural Networks, ESANN 2013, Bruges, Belgium, April 24-26, 2013.* URL: http://www.elen.ucl.ac.be/Proceedings/esann/esannpdf/es2013-123.pdf.

Kaufman, Kenton R. et al. (2008). “Energy Expenditure and Activity of Transfemoral Amputees Using Mechanical and Microprocessor-Controlled Prosthetic Knees”. In: *Archives* *of Physical Medicine and Rehabilitation* 89.7, pp. 1380–1385. ISSN: 0003-9993. DOI: 10.1016/j.apmr.2007.11.053. URL: https://dx.doi.org/10.1016/j.apmr.2007.11.053.

Kaushik, Saurav (2016). *Introduction to Feature Selection methods with an example (or how to select the right variables?)* Accessed: 2021-09. URL: https://www.analyticsvidhya.com/blog/2016/12/introduction-to-feature-selection-methods-with-an-example-or-how-to-select-the-right-variables/.

Keng, Brian (2016). *The Expectation-Maximization Algorithm.* Accessed: 2021-09. URL: https://bjlkeng.github.io/posts/the-expectation-maximization-algorithm/.

Keras (n.d.). Accessed: 2021-09. URL: https://keras.io/.

Khan, A M et al. (2010). "Human Activity Recognition via an Accelerometer-Enabled-Smartphone Using Kernel Discriminant Analysis". In: *2010 5th International Conference on Future Information Technology.* IEEE. DOI: 10.1109/futuretech.2010.5482729. URL: http://10.0.4.85/futuretech.2010.5482729%20https://dx.doi.org/10.1109/FUTURETECH.2010.5482729.

Khan, Adil, Y Lee, and Tae-Hun Kim (2008). "Accelerometer Signal-based Human Activity Recognition Using Augmented Autoregressive Model Coefficients and Artificial Neural Nets". In: *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference* 2008, pp. 5172–5. DOI: 10.1109/IEMBS.2008.4650379.

Khan, Adil, Muhammad Siddiqi, and Seok-Won Lee (2013). "Exploratory Data Analysis of Acceleration Signals to Select Light-Weight and Accurate Features for Real-Time Activity Recognition on Smartphones". In: *Sensors* 13.10, pp. 13099–13122. ISSN: 1424-8220. DOI: 10.3390/s131013099. URL: http://10.0.13.62/s131013099%20https://dx.doi.org/10.3390/s131013099.

Khan, Aftab et al. (2016). "Optimising sampling rates for accelerometer-based human activity recognition". In: *Pattern Recognition Letters* 73, pp. 33–40. ISSN: 0167-8655. DOI: https://doi.org/10.1016/j.patrec.2016.01.001. URL: http://www.sciencedirect.com/science/article/pii/S0167865516000040.

Khowaja, Sunder Ali, Bernardo Nugroho Yahya, and Seok-Lyong Lee (2020). "CAPHAR: context-aware personalized human activity recognition using associative learning in smart environments". In: *Human-centric Computing and Information Sciences* 10.1. ISSN: 2192-1962. DOI: 10.1186/s13673-020-00240-y.

Kimel-Naor, Shani, Amihai Gottlieb, and Meir Plotnik (2017). "The effect of uphill and downhill walking on gait parameters: A self-paced treadmill study". In: *Journal of Biomechanics* 60, pp. 142–149. ISSN: 0021-9290. DOI: 10.1016/j.jbiomech.2017.06.030. URL: http://10.0.3.248/j.jbiomech.2017.06.030%20https://dx.doi.org/10.1016/j.jbiomech.2017.06.030.

Kingma, Diederik P. and Jimmy Ba (2015). "Adam: A Method for Stochastic Optimization". In: *CoRR* abs/1412.6980.

Kitayama, Masaki (2021). *MATLAB-Kernel-PCA*. Accessed: 2021-09. URL: https://github.com/kitayama1234/MATLAB-Kernel-PCA.

Klute, Glenn et al. (2009). "Lower-limb amputee needs assessment using multistakeholder focus-group approach". In: *Journal of rehabilitation research and development* 46, pp. 293–304. DOI: 10.1682/JRRD.2008.02.0031.

Klute, Glenn K. et al. (2006). "Prosthetic Intervention Effects on Activity of Lower-Extremity Amputees". In: *Archives of Physical Medicine and Rehabilitation* 87.5, pp. 717–722. ISSN: 0003-9993. DOI: 10.1016/j.apmr.2006.02.007. URL: https://dx.doi.org/10.1016/j.apmr.2006.02.007.

Kononova, Anastasia et al. (2019). "The Use of Wearable Activity Trackers Among Older Adults: Focus Group Study of Tracker Perceptions, Motivators, and Barriers in the Maintenance Stage of Behavior Change". In: *JMIR mHealth and uHealth* 7.4, e9832. ISSN: 2291-5222. DOI: 10.2196/mhealth.9832. URL: http://10.0.8.148/mhealth.9832%20https://dx.doi.org/10.2196/mhealth.9832.

Koo, Terry K and Mae Y Li (2016). "A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research". In: *Journal of Chiropractic Medicine* 15.2, pp. 155–163. ISSN: 1556-3707. DOI: 10.1016/j.jcm.2016.02.012. URL: http://10.0.3.248/j.jcm.2016.02.012%20https://dx.doi.org/10.1016/j.jcm.2016.02.012.

Kosma, M., B. J. Cardinal, and J. A. McCubbin (2005). "A pilot study of a web-based physical activity motivational program for adults with physical disabilities". In: *Disabil Rehabil* 27.23, pp. 1435–42. ISSN: 0963-8288 (Print) 0963-8288. DOI: 10.1080/09638280500242713.

Krueger, R.A. (2014). *Focus Groups: A Practical Guide for Applied Research*. SAGE Publications, pp. 137–161. ISBN: 9781483365220. URL: https://books.google.co.uk/books?id=APtDBAAAQBAJ.

Kunze, Kai and Paul Lukowicz (2014). "Sensor Placement Variations in Wearable Activity Recognition". In: *IEEE Pervasive Computing* 13.4, pp. 32–41. ISSN: 1536-1268. DOI: 10.1109/mprv.2014.73. URL: http://10.0.4.85/mprv.2014.73%20https://dx.doi.org/10.1109/MPRV.2014.73.

Kwapisz, Jennifer R, Gary M Weiss, and Samuel A Moore (2011). "Activity recognition using cell phone accelerometers". In: *ACM SIGKDD Explorations Newsletter* 12.2, pp. 74–82. ISSN: 1931-0145. DOI: 10.1145/1964897.1964918. URL: http://10.0.4.121/1964897.1964918%20https://dx.doi.org/10.1145/1964897.1964918.

Kwon, Hyeokhyen, Gregory D Abowd, and Thomas Plötz (2019). "Handling annotation uncertainty in human activity recognition". In: *Proceedings of the 23rd International*

Symposium on Wearable Computers. ACM. DOI: 10.1145/3341163.3347744. URL: http://10.0.4.121/3341163.3347744%20https://dx.doi.org/10.1145/3341163.3347744.

Kwon, Yongjin, Kyuchang Kang, and Changseok Bae (2014). "Unsupervised learning for human activity recognition using smartphone sensors". In: *Expert Systems with Applications* 41.14, pp. 6067–6074. ISSN: 0957-4174. DOI: 10.1016/j.eswa.2014.04.037. URL: http://10.0.3.248/j.eswa.2014.04.037%20https://dx.doi.org/10.1016/j.eswa.2014.04.037.

Labarrière, Floriant et al. (2020). "Machine Learning Approaches for Activity Recognition and/or Activity Prediction in Locomotion Assistive Devices—A Systematic Review". In: *Sensors* 20.21. ISSN: 1424-8220. DOI: 10.3390/s20216345. URL: https://www.mdpi.com/1424-8220/20/21/6345.

Lai, B. et al. (2017). "Current Trends in Exercise Intervention Research, Technology, and Behavioral Change Strategies for People With Disabilities: A Scoping Review". In: *Am J Phys Med Rehabil* 96.10, pp. 748–761. ISSN: 0894-9115. DOI: 10.1097/phm.0000000000000743.

Lambert, Jeffrey D. et al. (2018). "Web-Based Intervention Using Behavioral Activation and Physical Activity for Adults With Depression (The eMotion Study): Pilot Randomized Controlled Trial". In: *J Med Internet Res* 20.7, e10112. ISSN: 1438-8871. DOI: 10.2196/10112. URL: http://www.jmir.org/2018/7/e10112/%20https://doi.org/10.2196/10112%20http://www.ncbi.nlm.nih.gov/pubmed/30012547.

Langford, J. et al. (2019). "Physical activity participation amongst individuals with lower limb amputation". In: *Disabil Rehabil* 41.9, pp. 1063–1070. ISSN: 0963-8288. DOI: 10.1080/09638288.2017.1422031.

LaPlante, M. (2016). *Amputation is no longer a surgery of last resort*. URL: https://www.deseret.com/2016/3/11/20584373/amputation-is-no-longer-a-surgery-of-last-resort#tyler-burdick-in-nevadas-ruby-mountains-in-late-january-2016.

Larkin, Louise et al. (2016). "Criterion Validity of the activPAL Activity Monitor for Sedentary and Physical Activity Patterns in People Who Have Rheumatoid Arthritis". In: *Physical Therapy* 96.7, pp. 1093–1101. ISSN: 0031-9023. DOI: 10.2522/ptj.20150281. URL: http://10.0.9.218/ptj.20150281%20https://dx.doi.org/10.2522/ptj.20150281.

Larson, Erika Rae (2015). "Massage therapy effects in a long-term prosthetic user with fibular hemimelia". In: *Journal of Bodywork and Movement Therapies* 19.2, pp. 261–267. ISSN: 1360-8592. DOI: 10.1016/j.jbmt.2014.04.005. URL: https://dx.doi.org/10.1016/j.jbmt.2014.04.005.

Lederman, Linda Costigan (1990). "Assessing educational effectiveness: The focus group interview as a technique for data collection". In: *Communication Education* 39.2, pp. 117–127. ISSN: 0363-4523. DOI: 10.1080/03634529009378794. URL: https://doi.org/10.1080/03634529009378794.

Lee, Heng-Ju and Li-Shan Chou (2007). “Balance control during stair negotiation in older adults”. In: *Journal of Biomechanics* 40.11, pp. 2530–2536. ISSN: 0021-9290. DOI: 10.1016/j.jbiomech.2006.11.001. URL: http://10.0.3.248/j.jbiomech.2006.11.001%20https://dx.doi.org/10.1016/j.jbiomech.2006.11.001.

Lee, M. S. et al. (2016). “Hand Gesture Recognition with Inertial Sensors and a Magnetometer”. In: *Sensors and Materials* 28, pp. 655–660. DOI: 10.18494/SAM.2016.1221.

Lee, Song-Mi, Sang Min Yoon, and Heeryon Cho (2017). “Human activity recognition from accelerometer data using Convolutional Neural Network”. In: *2017 IEEE International Conference on Big Data and Smart Computing (BigComp)*, pp. 131–134. DOI: 10.1109/BIGCOMP.2017.7881728.

Lee, Wangjoo, Seung-Hyeon Hong, and Hyun-Woo Oh (2018). “Characterization of Elastic Polymer-Based Smart Insole and a Simple Foot Plantar Pressure Visualization Method Using 16 Electrodes”. In: *Sensors* 19.1, p. 44. ISSN: 1424-8220. DOI: 10.3390/s19010044. URL: https://dx.doi.org/10.3390/s19010044.

Legro, Marcia et al. (1998). “Prosthesis evaluation questionnaire for persons with lower limb amputations: Assessing prosthesis-related quality of life”. In: *Archives of physical medicine and rehabilitation* 79, pp. 931–8. DOI: 10.1016/S0003-9993(98)90090-9.

Lemaire, Edward D and F.Ronald Fisher (1994). “Osteoarthritis and elderly amputee gait”. In: *Archives of Physical Medicine and Rehabilitation* 75.10, pp. 1094–1099. ISSN: 0003-9993. DOI: https://doi.org/10.1016/0003-9993(94)90084-1. URL: https://www.sciencedirect.com/science/article/pii/0003999394900841.

Leo, Marco, Tiziana D’Orazio, and Paolo Spagnolo (2004). “Human Activity Recognition for Automatic Visual Surveillance of Wide Areas”. In: *Proceedings of the ACM 2nd International Workshop on Video Surveillance; Sensor Networks*. VSSN ’04. New York, NY, USA: Association for Computing Machinery, pp. 124–130. ISBN: 1581139349. DOI: 10.1145/1026799.1026820.

Lester, Jonathan, Tanzeem Choudhury, and Gaetano Borriello (2006). “A Practical Approach to Recognizing Physical Activities”. In: *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, pp. 1–16. DOI: 10.1007/11748625_1. URL: http://10.0.3.239/11748625%7B%5C_%7D1%20https://dx.doi.org/10.1007/11748625%7B%5C_%7D1.

Lewis, Ann (1992). “Group Child Interviews as a Research Tool”. In: *British Educational Research Journal* 18.4, pp. 413–421. ISSN: 0141-1926. DOI: 10.1080/0141192920180407. URL: http://10.0.4.56/0141192920180407%20https://dx.doi.org/10.1080/0141192920180407.

Li, Shutong et al. (2019). “Accelerometer-Based Gyroscope Drift Compensation Approach in a Dual-Axial Stabilization Platform”. In: *Electronics* 8.5, p. 594. ISSN: 2079-9292. DOI: 10.3390/electronics8050594. URL: https://dx.doi.org/10.3390/electronics8050594.

Lin, Jessica et al. (2007). "Experiencing SAX: a novel symbolic representation of time series". In: *Data Mining and Knowledge Discovery* 15.2, pp. 107–144. ISSN: 1384-5810. DOI: 10.1007/s10618-007-0064-z. URL: http://10.0.3.239/s10618-007-0064-z%20https://dx.doi.org/10.1007/s10618-007-0064-z.

Lin, Weiyao et al. (June 2008). "Human activity recognition for video surveillance". In: pp. 2737–2740. DOI: 10.1109/ISCAS.2008.4542023.

Littman, A. J., E. D. Bouldin, and J. K. Haselkorn (2017). "This is your new normal: A qualitative study of barriers and facilitators to physical activity in Veterans with lower extremity loss". In: *Disabil Health J* 10.4, pp. 600–606. ISSN: 1876-7583 (Electronic) 1876-7583 (Linking). DOI: 10.1016/j.dhjo.2017.03.004. URL: https://www.ncbi.nlm.nih.gov/pubmed/28377115.

Littman, A. J. et al. (2014). "Physical activity barriers and enablers in older Veterans with lower-limb amputation". In: *J Rehabil Res Dev* 51.6, pp. 895–906. ISSN: 0748-7711. DOI: 10.1682/jrrd.2013.06.0152.

Littman, A. J. et al. (2019). "Pilot randomized trial of a telephone-delivered physical activity and weight management intervention for individuals with lower extremity amputation". In: *Disability and Health Journal* 12.1, pp. 43–50. ISSN: 1936-6574. DOI: 10.1016/j.dhjo.2018.08.002. URL: https://dx.doi.org/10.1016/j.dhjo.2018.08.002.

Liu, Hui and Tanja Schultz (2019). *A Wearable Real-time Human Activity Recognition System using Biosensors Integrated into a Knee Bandage*, pp. 47–55. DOI: 10.5220/0007398800470055.

Liu, Shaopeng et al. (2011). "SVM-based multi-sensor fusion for free-living physical activity assessment". In: *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE. DOI: 10.1109/iembs.2011.6090868. URL: http://10.0.4.85/iembs.2011.6090868%20https://dx.doi.org/10.1109/iembs.2011.6090868.

Logan, Beth (2000). "Mel Frequency Cepstral Coefficients for Music Modeling". In: *Proc. 1st Int. Symposium Music Information Retrieval*.

Logan, Beth et al. (2007). "A Long-Term Evaluation of Sensing Modalities for Activity Recognition". In: *UbiComp 2007: Ubiquitous Computing*. Ed. by John Krumm et al. Springer Berlin Heidelberg, pp. 483–500. ISBN: 978-3-540-74853-3.

Loiret, Isabelle et al. (2019). "Are wearable insoles a validated tool for quantifying transfemoral amputee gait asymmetry?" In: *Prosthetics & Orthotics International* 43.5, pp. 492–499. ISSN: 0309-3646. DOI: 10.1177/0309364619865814.

Lopez-Meyer, P., G. D. Fulk, and E. S. Sazonov (2011). "Automatic Detection of Temporal Gait Parameters in Poststroke Individuals". In: *IEEE Transactions on Information Technology in Biomedicine* 15.4, pp. 594–601. ISSN: 1089-7771. DOI: 10.1109/titb.2011.2112773.

Loukas, Serafeim (2020). *PCA clearly explained —When, Why, How to use it and feature importance: A guide in Python*. Accessed: 2022-02. URL: https://towardsdatascience.com/pca-clearly-explained-how-when-why-to-use-it-and-feature-importance-a-guide-in-python-7c274582c37e.

Lura, Derek J et al. (2017). "Crossover study of amputee stair ascent and descent biomechanics using Genium and C-Leg prostheses with comparison to non-amputee control". In: *Gait & Posture* 58, pp. 103–107. ISSN: 0966-6362. DOI: 10.1016/j.gaitpost.2017.07.114. URL: http://10.0.3.248/j.gaitpost.2017.07.114%20https://dx.doi.org/10.1016/j.gaitpost.2017.07.114.

Lustrek, Mitja, Bozidara Cvetkovic, and Simon Kozina (2012). "Energy expenditure estimation with wearable accelerometers". In: *2012 IEEE International Symposium on Circuits and Systems*. IEEE. DOI: 10.1109/iscas.2012.6271906. URL: http://10.0.4.85/iscas.2012.6271906%20https://dx.doi.org/10.1109/iscas.2012.6271906.

Ma, Haojie et al. (2021). "Unsupervised Human Activity Representation Learning with Multi-task Deep Clustering". In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5.1, pp. 1–25. ISSN: 2474-9567. DOI: 10.1145/3448074. URL: http://10.0.4.121/3448074%20https://dx.doi.org/10.1145/3448074.

Ma, Jasmin K. and Kathleen A. Martin Ginis (2018). "A meta-analysis of physical activity interventions in people with physical disabilities: Content, characteristics, and effects on behaviour". In: *Psychology of Sport and Exercise* 37, pp. 262–273. ISSN: 1469-0292. DOI: 10.1016/j.psychsport.2018.01.006. URL: https://dx.doi.org/10.1016/j.psychsport.2018.01.006.

Maaten, Laurens van der (2009). "Learning a Parametric Embedding by Preserving Local Structure". In: *Proceedings of the Twelth International Conference on Artificial Intelligence and Statistics*. Vol. 5. Proceedings of Machine Learning Research, pp. 384–391. URL: http://proceedings.mlr.press.

Maaten, Laurens van der and Geoffrey Hinton (2008). "Viualizing data using t-SNE". In: *Journal of Machine Learning Research* 9, pp. 2579–2605.

Machado, Inês P et al. (2015). "Human activity data discovery from triaxial accelerometer sensor: Non-supervised learning sensitivity to feature extraction parametrization". In: *Information Processing & Management* 51.2, pp. 204–214. ISSN: 0306-4573. DOI: 10.1016/j.ipm.2014.07.008. URL: http://10.0.3.248/j.ipm.2014.07.008%20https://dx.doi.org/10.1016/j.ipm.2014.07.008.

Mai, Jingeng et al. (2018). "Implementing a SoC-FPGA Based Acceleration System for On-Board SVM Training for Robotic Transtibial Prostheses". In: *2018 IEEE International Conference on Real-time Computing and Robotics (RCAR)*. IEEE. DOI: 10.1109/rcar.2018.8621732. URL: http://10.0.4.85/rcar.2018.8621732%20https://dx.doi.org/10.1109/rcar.2018.8621732.

Manda, P., G. Devi, and V. Row (2017). “A Random Forest based Classification Model for Human Activity Recognition”. In: *International Conference on Innovative Applications in Engineering and Information Technology*. Vol. 3. International Journal of Advanced Scientific Technologies, pp. 294–300.

Manini, A and A Sabatini (2010). “Machine learning methods for classifying human physical activity from on-body accelerometers.” In: *Sensors (Basel, Switzerland)* 10.2, pp. 1154–1175. ISSN: 1424-8220 (Electronic). DOI: 10.3390/s100201154.

Manini, T. M. and M. Pahor (2008). “Physical activity and maintaining physical function in older adults”. In: *British Journal of Sports Medicine* 43.1, pp. 28–31. ISSN: 0306-3674. DOI: 10.1136/bjsm.2008.053736. URL: https://dx.doi.org/10.1136/bjsm.2008.053736.

Mantyjarvi, J., J. Himberg, and T. Seppanen (2001). “Recognizing human motion with multiple acceleration sensors”. In: *2001 IEEE International Conference on Systems, Man and Cybernetics. e-Systems and e-Man for Cybernetics in Cyberspace (Cat.No.01CH37236)*. Vol. 2, 747–752 vol.2. DOI: 10.1109/ICSMC.2001.973004.

Maqbool, Hafiz Farhan et al. (2015). “Towards Intelligent Lower Limb Prostheses with Activity Recognition BT - Towards Autonomous Robotic Systems”. In: ed. by Clare Dixon and Karl Tuyls. Cham: Springer International Publishing, pp. 180–185. ISBN: 978-3-319-22416-9.

Marcot, Bruce G and Anca M Hanea (2021). “What is an optimal value of k in k-fold cross-validation in discrete Bayesian network analysis?” In: *Computational Statistics* 36.3, pp. 2009–2031. ISSN: 0943-4062. DOI: 10.1007/s00180-020-00999-9. URL: http://10.0.3.239/s00180-020-00999-9%20https://dx.doi.org/10.1007/s00180-020-00999-9.

Margoob, Mushtaq et al. (2008). “Prevalence of Post Traumatic Stress Disorder after Amputation: A Preliminary Study from Kashmir”. In: *JK practitioner: a journal of current clinical medicine & surgery* 15, pp. 5–7.

Martin, Bryan et al. (2017). “Methods for Real-Time Prediction of the Mode of Travel Using Smartphone-Based GPS and Accelerometer Data”. In: *Sensors* 17.9, p. 2058. ISSN: 1424-8220. DOI: 10.3390/s17092058. URL: https://dx.doi.org/10.3390/s17092058.

Mathie, M J et al. (2003). “Detection of daily physical activities using a triaxial accelerometer”. In: *Medical & Biological Engineering & Computing* 41.3, pp. 296–301. ISSN: 0140-0118. DOI: 10.1007/bf02348434. URL: http://10.0.3.239/bf02348434%20https://dx.doi.org/10.1007/bf02348434.

Mathworks (n.d.[a]). *Bayesian Optimization Workflow*. Accessed: 2021-09. URL: https://uk.mathworks.com/help/stats/bayesian-optimization-workflow.html.

— (n.d.[b]). *fscmrmr*. Accessed: 2021-09. URL: https://uk.mathworks.com/help/stats/fscmrmr.html.

Mathworks (n.d.[c]). *Hierarchical Clustering*. Accessed: 2021-09. URL: https://uk.mathworks.com/help/stats/hierarchical-clustering.html.

McGraw, Kenneth O and S P Wong (1996). "Forming inferences about some intraclass correlation coefficients." In: *Psychological Methods* 1.1, pp. 30–46. ISSN: 1939-1463(Electronic),1082-989X(Print). DOI: 10.1037/1082-989X.1.1.30.

McHugh, Mary L (2012). "Interrater reliability: the kappa statistic". In: *Biochemia medica* 22.3, pp. 276–282. ISSN: 1330-0962. URL: https://pubmed.ncbi.nlm.nih.gov/23092060%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3900052/.

McInnes, Leland, John Healy, and James Melville (2020). "UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction". In: *arXiv pre-print server*. DOI: None. URL: arxiv:1802.03426%20https://arxiv.org/abs/1802.03426%20file:///C:/Users/harpi/Downloads/arxiv-preprint-1802.03426%20(1).pdf.

McManus, Richard (2015). *Trackers: how technology is helping US Monitor and improve our health*. Bateman.

Md, Shah et al. (2016). *Human Activity Recognition using Smartphone Sensors with Context Filtering*.

Medicare (n.d.). *Local Coverage Determination (LCD): Lower Limb Prostheses (L33787). Centers for Medicare and Medicaid Services*. Accessed: 2021-09. URL: https://www.cms.gov/medicare-coverage-database./details/lcd-details.aspx?LCDId=33787%5C&ContrId=140%5C&ver=9%5C&ContrVer=2&CntrctrSelected.=140*2%5C&Cntrctr=140%5C&name=CGS+Administrators%5C%2c+LLC+(18003%5C%2c+DME+MAC)%5C&DocType=Active%5C&LCntrctr=140*2%5C&bc=AgACAAQAAAAAAA%5C%3d%5C%3d%5C&.

Meehan, Connor et al. (n.d.). *Uniform Manifold Approximation and Projection (UMAP)*. Accessed: 2021-09. URL: https://uk.mathworks.com/matlabcentral/fileexchange/71902-uniform-manifold-approximation-and-projection-umap.

Meg (2021). *Elevation on Strava FAQs*. Accessed: 2021-09. URL: https://support.strava.com/hc/en-us/articles/115001294564-Elevation-on-Strava-FAQs.

Metti, Andrea L. et al. (2018). "Longitudinal changes in physical function and physical activity in older adults". In: *Age and Ageing*. ISSN: 0002-0729. DOI: 10.1093/ageing/afy025. URL: https://dx.doi.org/10.1093/ageing/afy025.

Michie, Susan et al. (2011). "A refined taxonomy of behaviour change techniques to help people change their physical activity and healthy eating behaviours: The CALO-RE taxonomy". In: *Psychology & Health* 26.11, pp. 1479–1498. ISSN: 0887-0446. DOI: 10.1080/08870446.2010.540664. URL: https://dx.doi.org/10.1080/08870446.2010.540664.

Milenkoski, Martin et al. (2018). "Real time human activity recognition on smartphones using LSTM networks". In: *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, pp. 1126–1131. DOI: 10.23919/MIPRO.2018.8400205.

Miller, Matthew J. et al. (2017). "Physical activity behavior change for older veterans after dysvascular amputation". In: *Contemporary Clinical Trials* 55, pp. 10–15. ISSN: 1551-7144. DOI: 10.1016/j.cct.2017.01.008. URL: https://dx.doi.org/10.1016/j.cct.2017.01.008.

Miller, Matthew J. et al. (2019). "Factors influencing participation in physical activity after dysvascular amputation: a qualitative meta-synthesis". In: *Disability and Rehabilitation* 41.26, pp. 3141–3150. ISSN: 0963-8288. DOI: 10.1080/09638288.2018.1492031. URL: http://10.0.4.56/09638288.2018.1492031%20https://dx.doi.org/10.1080/09638288.2018.1492031.

mitosis (2020). *Support Vector Machine*. Accessed: 2021-09. URL: https://www.mitosistech.com/support-vector-machine/.

Mitzner, Tracy L et al. (2019). "Technology Adoption by Older Adults: Findings From the PRISM Trial". In: *The Gerontologist* 59.1, pp. 34–44. ISSN: 0016-9013. DOI: 10.1093/geront/gny113. URL: https://dx.doi.org/10.1093/geront/gny113.

Mohamad, Saad, Abdelhamid Bouchachia, and Moamar Sayed-Mouchaweh (2015). "A non-parametric hierarchical clustering model". In: *2015 IEEE International Conference on Evolving and Adaptive Intelligent Systems (EAIS)*. IEEE. DOI: 10.1109/eais.2015.7368803. URL: http://10.0.4.85/eais.2015.7368803%20https://dx.doi.org/10.1109/eais.2015.7368803.

Molina, L.C., L. Belanche, and A. Nebot (n.d.). "Feature selection algorithms: a survey and experimental evaluation". In: DOI: 10.1109/icdm.2002.1183917.

Montesinos-Magraner, L. et al. (2016). "Physical and Psychosocial Functions of Adults with Lower Limb Congenital Deficiencies and Amputations in Childhood". In: *Rehabilitation Research and Practice* 2016, pp. 1–7. ISSN: 2090-2867. DOI: 10.1155/2016/8109365. URL: https://dx.doi.org/10.1155/2016/8109365.

Morelhão, Priscila, Crystian Oliveira, and Marcia Franco (2016). "Interventions to increase physical activity among older adults (PEDro synthesis)". In: *British Journal of Sports Medicine* 51, bjsports–2016. DOI: 10.1136/bjsports-2016-096859.

Muhammad Masum, Abdul Kadar et al. (2018). "Human Activity Recognition Using Multiple Smartphone Sensors". In: *2018 International Conference on Innovations in Science, Engineering and Technology (ICISET)*. IEEE. DOI: 10.1109/iciset.2018.8745628. URL: http://10.0.4.85/iciset.2018.8745628%20https://dx.doi.org/10.1109/ICISET.2018.8745628.

Muzaffar, N. et al. (2012). "Psychiatric comorbidity in amputees with average sociodemographic status and the role of theologic and family support in a conflict zone". In: *Australasian Journal of Disaster and Trauma Studies* 2012, pp. 31–38.

Najafi, B et al. (2003). "Ambulatory system for human motion analysis using a kinematic sensor: monitoring of daily physical activity in the elderly". In: *IEEE Transactions on Biomedical Engineering* 50.6, pp. 711–723. ISSN: 0018-9294. DOI: 10.1109/tbme.2003.812189. URL: http://10.0.4.85/tbme.2003.812189%20https://dx.doi.org/10.1109/TBME.2003.812189.

Nam, Yunyoung and Jung Wook Park (2013). "Child Activity Recognition Based on Cooperative Fusion Model of a Triaxial Accelerometer and a Barometric Pressure Sensor". In: *IEEE Journal of Biomedical and Health Informatics* 17.2, pp. 420–426. ISSN: 2168-2194. DOI: 10.1109/jbhi.2012.2235075. URL: http://10.0.4.85/jbhi.2012.2235075%20https://dx.doi.org/10.1109/JBHI.2012.2235075.

Narang, I C et al. (1984). "Functional capabilities of lower limb amputees". In: *Prosthetics & Orthotics International* 8.1, pp. 43–51. ISSN: 0309-3646. DOI: 10.3109/03093648409145345. URL: http://10.0.12.37/03093648409145345%20https://dx.doi.org/10.3109/03093648409145345.

Ng, Andrew (n.d.). *Stanford Machine Learning Online Course*. Accessed: 2021-09. URL: https://www.coursera.org/learn/machine-learning.

Ngueleu, Armelle M. et al. (2019). "Validity of Instrumented Insoles for Step Counting, Posture and Activity Recognition: A Systematic Review". In: *Sensors* 19.11, p. 2438. ISSN: 1424-8220. DOI: 10.3390/s19112438. URL: https://dx.doi.org/10.3390/s19112438.

Nguyen, Le Nguyen Ngu et al. (2015). "Basketball Activity Recognition using Wearable Inertial Measurement Units". In: *Proceedings of the XVI International Conference on Human Computer Interaction*. ACM. DOI: 10.1145/2829875.2829930. URL: http://10.0.4.121/2829875.2829930%20https://dx.doi.org/10.1145/2829875.2829930.

Nguyen, Nhan Duc et al. (2018). "Position-Based Feature Selection for Body Sensors regarding Daily Living Activity Recognition". In: *Journal of Sensors* 2018. Ed. by Eduard Llobet, p. 9762098. ISSN: 1687-725X. DOI: 10.1155/2018/9762098. URL: https://doi.org/10.1155/2018/9762098.

Nguyen, Truc D T, Trung-Tin Huynh, and Hoang-Anh Pham (2018). "An Improved Human Activity Recognition by Using Genetic Algorithm to Optimize Feature Vector". In: *2018 10th International Conference on Knowledge and Systems Engineering (KSE)*. IEEE. DOI: 10.1109/kse.2018.8573335. URL: http://10.0.4.85/kse.2018.8573335%20https://dx.doi.org/10.1109/kse.2018.8573335.

NHS Information Centre (2011). "National Diabetes Audit Executive Summary 2009 - 2010". In: *National Health Service*.

Ni, Qin, Lei Zhang, and Luqun Li (2018). “A Heterogeneous Ensemble Approach for Activity Recognition with Integration of Change Point-Based Data Segmentation”. In: *Applied Sciences* 8.9, p. 1695. ISSN: 2076-3417. DOI: 10.3390/app8091695. URL: http://10.0.13.62/app8091695%20https://dx.doi.org/10.3390/app8091695.

Nistler, Jonathan R and Majura F Selekwa (2011). “Gravity compensation in accelerometer measurements for robot navigation on inclined surfaces”. In: *Procedia Computer Science* 6, pp. 413–418. ISSN: 1877-0509. DOI: 10.1016/j.procs.2011.08.077. URL: http://10.0.3.248/j.procs.2011.08.077%20https://dx.doi.org/10.1016/j.procs.2011.08.077.

Novick, Gina (2008). “Is there a bias against telephone interviews in qualitative research?” In: *Research in Nursing & Health* 31.4, pp. 391–398. ISSN: 0160-6891. DOI: 10.1002/nur.20259. URL: http://10.0.3.234/nur.20259%20https://dx.doi.org/10.1002/nur.20259.

Nyan, M N et al. (2006). “Classification of gait patterns in the time–frequency domain”. In: *Journal of Biomechanics* 39.14, pp. 2647–2656. ISSN: 0021-9290. DOI: https://doi.org/10.1016/j.jbiomech.2005.08.014. URL: https://www.sciencedirect.com/science/article/pii/S0021929005003891.

O’Brien, Ciara M et al. (2020). “Measurement of sedentary time and physical activity in rheumatoid arthritis: an ActiGraph and activPAL™ validation study”. In: *Rheumatology International* 40.9, pp. 1509–1518. ISSN: 0172-8172. DOI: 10.1007/s00296-020-04608-2. URL: http://10.0.3.239/s00296-020-04608-2%20https://dx.doi.org/10.1007/s00296-020-04608-2.

O’Brien, Megan et al. (2017). “Activity Recognition for Persons With Stroke Using Mobile Phone Technology: Toward Improved Performance in a Home Setting”. In: *Journal of Medical Internet Research* 19, e184. DOI: 10.2196/jmir.7385.

Olah, Christopher (2015). *Understanding LSTM Networks*. Accessed: 2021-09. URL: https://colah.github.io/posts/2015-08-Understanding-LSTMs/.

Omkar, Sn, Meenakshi Mour, and Debarun Das (2009). “Motion analysis of sun salutation using magnetometer and accelerometer”. In: *International journal of yoga* 2.2, pp. 62–68. ISSN: 0973-6131. DOI: 10.4103/0973-6131.60046. URL: http://europepmc.org/abstract/MED/20842266%20https://www.ncbi.nlm.nih.gov/pmc/articles/pmid/20842266/?tool=EBI%20https://doi.org/10.4103/0973-6131.60046%20https://europepmc.org/articles/PMC2933730.

Orendurff, Michael (2016). “Gait During Real-World Challenges: Gait Initiation, Gait Termination, Acceleration, Deceleration, Turning, Slopes, and Stairs”. In: *Handbook of Human Motion*. Springer International Publishing, pp. 1–21. DOI: 10.1007/978-3-319-30808-1_47-1. URL: http://10.0.3.239/978-3-319-30808-1%7B%5C_%7D47-1%20https://dx.doi.org/10.1007/978-3-319-30808-1%7B%5C_%7D47-1.

OttoBock (n.d.). *C-Leg 3C98-3/3C88-3 Manual*. URL: https://www.ottobockus.com/media/local-media/prosthetics/lower-limb/c-leg/files/cleg4-ifu.pdf.

Ozcan, Tayyip and Alper Basturk (2020). “Human action recognition with deep learning and structural optimization using a hybrid heuristic algorithm”. In: *Cluster Computing* 23.4, pp. 2847–2860. ISSN: 1386-7857. DOI: 10.1007/s10586-020-03050-0. URL: http://10.0.3.239/s10586-020-03050-0%20https://dx.doi.org/10.1007/s10586-020-03050-0.

PAL-Technologies (n.d.). *PAL family*. Accessed: 2021-09. URL: https://www.palt.com/pals/.

Pandeirot, Lanemey and Andrew Aseng (2017). “Social Loafing and Group Performance: A Literature Review”. In: *5th International Scholars Conference*.

Panwar, Madhuri et al. (2017). “CNN based approach for activity recognition using a wrist-worn accelerometer”. In: *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 2438–2441. DOI: 10.1109/EMBC.2017.8037349.

Park, Sang-Eun et al. (2019). “Measurement and Analysis of Gait Pattern during Stair Walk for Improvement of Robotic Locomotion Rehabilitation System”. In: *Applied Bionics and Biomechanics* 2019, pp. 1–12. ISSN: 1176-2322. DOI: 10.1155/2019/1495289. URL: http://10.0.4.131/2019/1495289%20https://dx.doi.org/10.1155/2019/1495289.

Pasquina, P. F, A. J. Carvalho, and T. P. Sheehan (2015). “Ethics in Rehabilitation: Access to Prosthetics and Quality Care Following Amputation”. In: *AMA Journal of Ethics* 17.6, pp. 535–546. ISSN: 2376-6980. DOI: 10.1001/journalofethics.2015.17.6.stas1-1506. URL: https://dx.doi.org/10.1001/journalofethics.2015.17.6.stas1-1506.

Pasquina, Paul F. et al. (2015). “Recent advances in bioelectric prostheses”. In: *Neurology. Clinical practice* 5.2, pp. 164–170. ISSN: 2163-0402 2163-0933. DOI: 10.1212/CPJ.0000000000000132. URL: https://www.ncbi.nlm.nih.gov/pubmed/29443190%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5764448/.

Passero, Thomas (2014). “Devising the Prosthetic Prescription and Typical Examples”. In: *Physical Medicine and Rehabilitation Clinics of North America* 25.1, pp. 117–132. ISSN: 1047-9651. DOI: https://doi.org/10.1016/j.pmr.2013.09.009. URL: http://www.sciencedirect.com/science/article/pii/S1047965113000764.

Pavey, Toby G. et al. (2017). “Field evaluation of a random forest activity classifier for wrist-worn accelerometer data”. In: *Journal of Science and Medicine in Sport* 20.1, pp. 75–80. ISSN: 1440-2440. DOI: 10.1016/j.jsams.2016.06.003. URL: https://dx.doi.org/10.1016/j.jsams.2016.06.003.

Pear-Stairs (n.d.). *UK Staircase Building Regulations*. Accessed: 2022-02. URL: https://www.pearstairs.co.uk/staircase-buillding-regulations.

Pedley, Mark (2013). “Tilt Sensing Using a Three-Axis Accelerometer”. In: URL: https://www.nxp.com/files-static/sensors/doc/app_note/AN3461.pdf.

Peña, J M, J A Lozano, and P Larrañaga (1999). “An empirical comparison of four initialization methods for the K-Means algorithm”. In: *Pattern Recognition Letters* 20.10, pp. 1027–1040. ISSN: 0167-8655. DOI: 10.1016/s0167-8655(99)00069-0. URL: http://10.0.3.248/s0167-8655(99)00069-0%20https://dx.doi.org/10.1016/s0167-8655(99)00069-0.

Pepin, M. E., K. G. Akers, and S. S. Galen (2018). “Physical activity in individuals with lower extremity amputations: a narrative review”. In: *Physical Therapy Reviews* 23.2, pp. 77–87. ISSN: 1083-3196. DOI: 10.1080/10833196.2017.1412788. URL: https://doi.org/10.1080/10833196.2017.1412788.

Pepin, Marie et al. (2018). “Correlation Between Functional Ability and Physical Activity in Individuals With Transtibial Amputations: A Cross-Sectional Study”. In: *Cardiopulmonary Physical Therapy Journal* 30, p. 1. DOI: 10.1097/CPT.0000000000000091.

Perry, Jacquelin and Judith M Burnfield (2010). “Gait Analysis: Normal and Pathological Function”. In: *Journal of Sports Science & Medicine* 9.2, p. 353. ISSN: 1303-2968. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3761742/%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3761742/.

Pfister, Ted et al. (2017). “Comparison of two accelerometers for measuring physical activity and sedentary behaviour”. In: *BMJ Open Sport & Exercise Medicine* 3.1. ISSN: 2055-7647. DOI: 10.1136/bmjsem-2017-000227. URL: https://dx.doi.org/10.1136/bmjsem-2017-000227.

Piezon, Sherry L and William D Ferree (2008). “Perceptions of Social Loafing in Online Learning Groups: A study of Public University and U.S. Naval War College students”. In: *The International Review of Research in Open and Distributed Learning* 9.2. ISSN: 1492-3831. DOI: 10.19173/irrodl.v9i2.484. URL: http://10.0.74.229/irrodl.v9i2.484%20https://dx.doi.org/10.19173/irrodl.v9i2.484.

Ploeg, H. P. van der et al. (2006). “Counselling increases physical activity behaviour nine weeks after rehabilitation”. In: *Br J Sports Med* 40.3, pp. 223–9. ISSN: 0306-3674. DOI: 10.1136/bjsm.2005.021139.

Ploeg, H. P. van der et al. (2007). “The Physical Activity Scale for Individuals with Physical Disabilities: test-retest reliability and comparison with an accelerometer”. In: *J Phys Act Health* 4.1, pp. 96–100. ISSN: 1543-3080 (Print) 1543-3080.

Podsiadlo, D. and S. Richardson (1991). “The timed "Up & Go": a test of basic functional mobility for frail elderly persons”. In: *J Am Geriatr Soc* 39.2, pp. 142–8. ISSN: 0002-8614 (Print) 0002-8614. DOI: 10.1111/j.1532-5415.1991.tb01616.x.

Polar-Electro (n.d.). *Polar Grit X*. Accessed: 2022-02. URL: https://www.polar.com/uk-en/grit-x.

Powers, C et al. (1997). "Stair ambulation in persons with transtibial amputation: An analysis of the Seattle LightFoot(TM)". In: *Journal of rehabilitation research and development* 34, pp. 9–18.

Powers, C M, S Rao, and J Perry (1998). "Knee kinetics in trans-tibial amputee gait." In: *Gait & posture* 8.1, pp. 1–7. ISSN: 1879-2219 (Electronic). DOI: 10.1016/s0966-6362(98)00016-2.

Preece, Stephen J et al. (2009). "A Comparison of Feature Extraction Methods for the Classification of Dynamic Activities From Accelerometer Data". In: *IEEE Transactions on Biomedical Engineering* 56.3, pp. 871–879. ISSN: 0018-9294. DOI: 10.1109/tbme.2008.2006190. URL: http://10.0.4.85/tbme.2008.2006190%20https://dx.doi.org/10.1109/TBME.2008.2006190.

Protopapadaki, Anastasia et al. (2007). "Hip, knee, ankle kinematics and kinetics during stair ascent and descent in healthy young individuals". In: *Clinical biomechanics (Bristol, Avon)* 22, pp. 203–10. DOI: 10.1016/j.clinbiomech.2006.09.010.

Qazi, Nadeem and Kamran Raza (2012). "Effect of Feature Selection, SMOTE and under Sampling on Class Imbalance Classification". In: *2012 UKSim 14th International Conference on Computer Modelling and Simulation*. IEEE. DOI: 10.1109/uksim.2012.116. URL: http://10.0.4.85/uksim.2012.116%20https://dx.doi.org/10.1109/uksim.2012.116.

Qi, Jun et al. (2018). "Examining sensor-based physical activity recognition and monitoring for healthcare using Internet of Things: A systematic review". In: *Journal of Biomedical Informatics* 87, pp. 138–153. ISSN: 1532-0464. DOI: https://doi.org/10.1016/j.jbi.2018.09.002. URL: http://www.sciencedirect.com/science/article/pii/S153204641830176X.

Quirk, Helen, Cris Glazebrook, and Holly Blake (2018). "A physical activity intervention for children with type 1 diabetes- steps to active kids with diabetes (STAK-D): a feasibility study". In: *BMC Pediatrics* 18.1. ISSN: 1471-2431. DOI: 10.1186/s12887-018-1036-8. URL: https://dx.doi.org/10.1186/s12887-018-1036-8.

Raichle, Katherine A. et al. (2008). "Prosthesis use in persons with lower- and upper-limb amputation". In: *Journal of rehabilitation research and development* 45.7, pp. 961–972. ISSN: 1938-1352 0748-7711. DOI: 10.1682/jrrd.2007.09.0151. URL: https://pubmed.ncbi.nlm.nih.gov/19165686%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2743731/.

DC-Rainmaker (n.d.). *Polar Grit X GPS Watch In-Depth Review*. Accessed: 2022-02. URL: https://www.dcrainmaker.com/2020/04/polar-grit-x-gps-watch-review.html.

Raschka, Sebastian (n.d.). *Model evaluation, model selection, and algorithm selection in machine learning Part IV - Comparing the performance of machine learning models and algorithms using statistical tests and nested cross-validation*. Accessed: 2022-02. URL: https://sebastianraschka.com/blog/2018/model-evaluation-selection-part4.html.

Rasekh, Amin, Chien-An Chen, and Yan Lu (2014). *Human Activity Recognition using Smartphone*. Tech. rep. Texas A&M University. DOI: 10.35940/ijrte.d4521.118419.

Ravi, Nishkam et al. (2005). "Activity Recognition from Accelerometer Data." In: *AAAI*. Vol. 3, pp. 1541–1546.

Reiner, Miriam et al. (2013). "Long-term health benefits of physical activity – a systematic review of longitudinal studies". In: *BMC Public Health* 13.1, p. 813. ISSN: 1471-2458. DOI: 10.1186/1471-2458-13-813. URL: https://dx.doi.org/10.1186/1471-2458-13-813.

Resnik, L., M. Borgia, and B. Silver (2017). "Measuring Community Integration in Persons With Limb Trauma and Amputation: A Systematic Review". In: *Arch Phys Med Rehabil* 98.3, 561–580.e8. ISSN: 0003-9993. DOI: 10.1016/j.apmr.2016.08.463.

Resnik, Linda and Matthew Borgia (2011). "Reliability of Outcome Measures for People With Lower-Limb Amputations: Distinguishing True Change From Statistical Error". In: *Physical Therapy* 91.4, pp. 555–565. ISSN: 0031-9023. DOI: 10.2522/ptj.20100287. eprint: https://academic.oup.com/ptj/article-pdf/91/4/555/17503744/ptj0555.pdf. URL: https://doi.org/10.2522/ptj.20100287.

Review, World Population (n.d.). Accessed: 2022-01. URL: https://worldpopulationreview.com/country-rankings/military-size-by-country.

Riener, Robert, Marco Rabuffetti, and Carlo Frigo (2002). "Stair ascent and descent at different inclinations". In: *Gait & Posture* 15.1, pp. 32–44. ISSN: 0966-6362. DOI: 10.1016/s0966-6362(01)00162-x. URL: http://10.0.3.248/s0966-6362(01)00162-x%20https://dx.doi.org/10.1016/s0966-6362(01)00162-x.

Robertson, Neil and Ian Reid (2006). "A general method for human activity recognition in video". In: *Computer Vision and Image Understanding* 104.2-3, pp. 232–248. ISSN: 1077-3142. DOI: 10.1016/j.cviu.2006.07.006.

Rocca, Joesph (2019). *Ensemble methods: bagging, boosting and stacking.* Accessed: 2021-09. URL: https://towardsdatascience.com/ensemble-methods-bagging-boosting-and-stacking-c9214a10a205.

Rodrigues, Fábio Barbosa, Adriano O Andrade, and Marcus Fraga Vieira (2019). "Effects of inclined surfaces on gait variability and stability in unilateral lower limb amputees". In: *Medical & Biological Engineering & Computing* 57.11, pp. 2337–2346. ISSN: 0140-0118. DOI: 10.1007/s11517-019-02042-6. URL: http://10.0.3.239/s11517-019-02042-6%20https://dx.doi.org/10.1007/s11517-019-02042-6.

Roetenberg, D et al. (2005). "Compensation of magnetic disturbances improves inertial and magnetic sensing of human body segment orientation". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 13.3, pp. 395–405. ISSN: 1534-4320. DOI: 10.1109/tnsre.2005.847353. URL: http://10.0.4.85/tnsre.2005.847353%20https://dx.doi.org/10.1109/tnsre.2005.847353.

Romano, Simone et al. (2015). "Adjusting for Chance Clustering Comparison Measures". In: *arXiv pre-print server*. DOI: None. URL: arxiv:1512.01286%20https://arxiv.org/abs/1512.01286.

Rosati, Samanta, Gabriella Balestra, and Marco Knaflitz (2018). "Comparison of Different Sets of Features for Human Activity Recognition by Wearable Sensors". In: *Sensors* 18.12, p. 4189. ISSN: 1424-8220. DOI: 10.3390/s18124189. URL: http://10.0.13.62/s18124189%20https://dx.doi.org/10.3390/s18124189.

Rousseeuw, Peter J. (1987). "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis". In: *Journal of Computational and Applied Mathematics* 20, pp. 53–65. ISSN: 0377-0427. DOI: https://doi.org/10.1016/0377-0427(87)90125-7. URL: https://www.sciencedirect.com/science/article/pii/0377042787901257.

Russell, Brian et al. (2021). "Moving the Lab into the Mountains: A Pilot Study of Human Activity Recognition in Unstructured Environments". In: *Sensors* 21.2, p. 654. ISSN: 1424-8220. DOI: 10.3390/s21020654. URL: http://10.0.13.62/s21020654%20https://dx.doi.org/10.3390/s21020654.

Ryan, C G (2006). "The validity and reliability of a novel activity monitor as a measure of walking". In: *British Journal of Sports Medicine* 40.9, pp. 779–784. ISSN: 0306-3674. DOI: 10.1136/bjsm.2006.027276. URL: http://10.0.4.112/bjsm.2006.027276%20https://dx.doi.org/10.1136/bjsm.2006.027276.

Sabatini, A. M. et al. (2005). "Assessment of Walking Features From Foot Inertial Sensing". In: *IEEE Transactions on Biomedical Engineering* 52.3, pp. 486–494. ISSN: 0018-9294. DOI: 10.1109/tbme.2004.840727. URL: https://dx.doi.org/10.1109/TBME.2004.840727.

Sahu, Anamika et al. (2016). "Psychological effects of amputation: A review of studies from India". In: *Industrial psychiatry journal* 25.1, pp. 4–10. ISSN: 0972-6748 0976-2795. DOI: 10.4103/0972-6748.196041. URL: https://pubmed.ncbi.nlm.nih.gov/28163401%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5248418/.

Sainburg, Tim, Leland McInnes, and Timothy (2020). "Parametric UMAP: learning embeddings with deep neural networks for representation and semi-supervised learning". In: *arXiv pre-print server*. DOI: None. URL: arxiv:2009.12981%20https://arxiv.org/abs/2009.12981%20file:///C:/Users/harpi/Downloads/arxiv-preprint-2009.12981%20(1).pdf.

Saleme, C.J. et al. (2013). "Syme ankle disarticulation and transtibial amputation as treatments for pelvic limb deficiencies. Which is better?" In: *Rev Mex Ortop Ped* 15.2, pp. 79–84.

Salter, K. et al. (2005). "Issues for selection of outcome measures in stroke rehabilitation: ICF activity". In: *Disabil Rehabil* 27.6, pp. 315–40. ISSN: 0963-8288 (Print) 0963-8288. DOI: 10.1080/09638280400008545.

Samuelsen, Brian T. et al. (2017). “The Impact of the Immediate Postoperative Prosthesis on Patient Mobility and Quality of Life after Transtibial Amputation”. In: *American Journal of Physical Medicine & Rehabilitation* 96.2, pp. 116–119. ISSN: 0894-9115. DOI: 10.1097/phm.0000000000000553. URL: https://journals.lww.com/ajpmr/Fulltext/2017/02000/The_Impact_of_the_Immediate_Postoperative.8.aspx.

Samuelsson, Kersti Am et al. (2012). “Effects of lower limb prosthesis on activity, participation, and quality of life: a systematic review”. In: *Prosthetics and Orthotics International* 36.2, pp. 145–158. ISSN: 0309-3646. DOI: 10.1177/0309364611432794. URL: https://dx.doi.org/10.1177/0309364611432794.

San Buenaventura, Charlene V and Nestor Michael C Tiglao (2017). “Basic Human Activity Recognition based on sensor fusion in smartphones”. In: *2017 IFIP/IEEE Symposium on Integrated Network and Service Management (IM)*. IEEE. DOI: 10.23919/inm.2017.7987459. URL: http://10.0.93.111/inm.2017.7987459%20https://dx.doi.org/10.23919/INM.2017.7987459.

San-Segundo, Rubén et al. (2016). “Feature extraction from smartphone inertial signals for human activity segmentation”. In: *Signal Processing* 120, pp. 359–372. ISSN: 0165-1684. DOI: 10.1016/j.sigpro.2015.09.029. URL: http://10.0.3.248/j.sigpro.2015.09.029%20https://dx.doi.org/10.1016/j.sigpro.2015.09.029.

San-Segundo, Rubén et al. (2018). “Robust Human Activity Recognition using smartwatches and smartphones”. In: *Engineering Applications of Artificial Intelligence* 72, pp. 190–202. ISSN: 0952-1976. DOI: 10.1016/j.engappai.2018.04.002. URL: http://10.0.3.248/j.engappai.2018.04.002%20https://dx.doi.org/10.1016/j.engappai.2018.04.002.

Sanders, Joan et al. (2018). “A Novel Method for Assessing Prosthesis Use and Accommodation Practices of People with Transtibial Amputation”. In: *Journal of Prosthetics and Orthotics* 30, p. 1. DOI: 10.1097/JPO.0000000000000209.

Sani, Sadiq et al. (2017). “kNN sampling for personalised human recognition.” In: ed. by David W. Aha and Jean Lieber. Springer, pp. 330–344. ISBN: 9783319610290. DOI: 10.1007/978-3-319-61030-6_23. URL: http://hdl.handle.net/10059/2486.

Schmalz, Thomas, Siegmar Blumentritt, and Björn Marx (2007). “Biomechanical analysis of stair ambulation in lower limb amputees”. In: *Gait & Posture* 25.2, pp. 267–278. ISSN: 0966-6362. DOI: 10.1016/j.gaitpost.2006.04.008. URL: http://10.0.3.248/j.gaitpost.2006.04.008%20https://dx.doi.org/10.1016/j.gaitpost.2006.04.008.

Schmidt, Michael D. et al. (2009). “Cardiometabolic Risk in Younger and Older Adults Across an Index of Ambulatory Activity”. In: 37.4, pp. 278–284. ISSN: 0749-3797. DOI: 10.1016/j.amepre.2009.05.020. URL: https://dx.doi.org/10.1016/j.amepre.2009.05.020.

Schrader, Lisa et al. (2020). “Advanced Sensing and Human Activity Recognition in Early Intervention and Rehabilitation of Elderly People”. In: *Journal of Population Ageing*.

ISSN: 1874-7884. DOI: 10.1007/s12062-020-09260-z. URL: http://10.0.3.239/s12062-020-09260-z%20https://dx.doi.org/10.1007/s12062-020-09260-z.

scikit-learn (n.d.). *Importance of Feature Scaling*. Accessed: 2021-09. URL: https://scikit-learn.org/stable/auto_examples/preprocessing/plot_scaling_importance.html.

Scopes, J. et al. (2015). "The bacpar outcome measures toolbox: A step towards standardising outcome measures for physiotherapist working with lower limb amputees". In: *Physiotherapy (United Kingdom)* 101.SUPPL. 1, eS1357–eS1358. ISSN: 0031-9406. DOI: http://dx.doi.org/10.1016/j.physio.2015.03.1293.

Sekine, M et al. (2002). "Discrimination of walking patterns using wavelet-based fractal analysis". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 10.3, pp. 188–196. ISSN: 1534-4320. DOI: 10.1109/tnsre.2002.802879. URL: http://10.0.4.85/tnsre.2002.802879%20https://dx.doi.org/10.1109/TNSRE.2002.802879.

Sekine, Masaki et al. (2000). "Classification of waist-acceleration signals in a continuous walking record". In: *Medical Engineering & Physics* 22.4, pp. 285–291. ISSN: 1350-4533. DOI: 10.1016/s1350-4533(00)00041-2. URL: http://10.0.3.248/s1350-4533(00)00041-2%20https://dx.doi.org/10.1016/s1350-4533(00)00041-2.

Senra, Hugo et al. (2012). "Beyond the body image: a qualitative study on how adults experience lower limb amputation". In: *Clinical Rehabilitation* 26.2, pp. 180–191. ISSN: 0269-2155. DOI: 10.1177/0269215511410731. URL: http://10.0.4.153/0269215511410731%20https://dx.doi.org/10.1177/0269215511410731.

Shahriari, B et al. (2016). "Taking the Human Out of the Loop: A Review of Bayesian Optimization". In: *Proceedings of the IEEE* 104.1, pp. 148–175. ISSN: 1558-2256. DOI: 10.1109/JPROC.2015.2494218.

Shapiro, S S and M B Wilk (1965). "An Analysis of Variance Test for Normality (Complete Samples)". In: *Biometrika* 52.3/4, pp. 591–611. ISSN: 00063444. DOI: 10.2307/2333709. URL: http://www.jstor.org/stable/2333709.

Sharma, Sagar (2017). *Activation Functions in Neural Networks*. Accessed: 2021-09. URL: https://towardsdatascience.com/activation-functions-neural-networks-1cbd9f8d91d6.

Shawen, Nicholas et al. (2017). "Fall Detection in Individuals With Lower Limb Amputations Using Mobile Phones: Machine Learning Enhances Robustness for Real-World Applications." In: *JMIR mHealth and uHealth* 5.10, e151. ISSN: 2291-5222 (Print). DOI: 10.2196/mhealth.8201.

Shdefat, Ahmed, Ahmed Halimeh, and Hee Kim (2018). "Human Activities Recognition Via Smartphones Using Supervised Machine Learning Classifiers". In: *Primary Health Care Open Access* 08. DOI: 10.4172/2167-1079.1000289.

Shin, Min Kyung et al. (2018). “Effects of Lumbar Strengthening Exercise in Lower-Limb Amputees With Chronic Low Back Pain”. In: *Annals of Rehabilitation Medicine* 42.1, p. 59. ISSN: 2234-0645. DOI: 10.5535/arm.2018.42.1.59. URL: https://dx.doi.org/10.5535/arm.2018.42.1.59.

Shoaib, Muhammad et al. (2014). “Fusion of Smartphone Motion Sensors for Physical Activity Recognition”. In: *Sensors* 14.6, pp. 10146–10176. ISSN: 1424-8220. DOI: 10.3390/s140610146. URL: http://10.0.13.62/s140610146%20https://dx.doi.org/10.3390/s140610146.

Shrout, P E and J L Fleiss (1979). “Intraclass correlations: uses in assessing rater reliability.” In: *Psychological bulletin* 86.2, pp. 420–428. ISSN: 0033-2909 (Print). DOI: 10.1037//0033-2909.86.2.420.

Siirtola, Pekka, Heli Koskimäki, and Juha Röning (2019). *Personalizing human activity recognition models using incremental learning*. eprint: 1905.12628.

Silva, Felipe O, Lucas P S Paiva, and Gustavo S Carvalho (2021). “Error Analysis of Accelerometer- and Magnetometer-Based Stationary Alignment”. In: *Sensors* 21.6. DOI: 10.3390/s21062040.

SketchUP3DConstruction (n.d.). *Definition, Types and Benefits of road cambers in highway*. Accessed: 2021-09. URL: http://www.sketchup3dconstruction.com/const/types-and-benefits-of-road-cambers-in-highway.html.

Smith, John, Gary Guerra, and Brian Burkholder (2019). “The validity and accuracy of wrist-worn activity monitors in lower-limb prosthesis users”. In: *Disability and Rehabilitation*, pp. 1–7. DOI: 10.1080/09638288.2019.1587792.

Song, Kai-Tai and Yao-qing Wang (2005). “Remote Activity Monitoring of the Elderly Using a Two-Axis Accelerometer”. In: *Proceedings of theCACS Automatic Control Conference*.

Sprint, Gina, Diane J. Cook, and Douglas L. Weeks (2015). “Toward Automating Clinical Assessments: A Survey of the Timed Up and Go”. In: *IEEE reviews in biomedical engineering* 8, pp. 64–77. ISSN: 1941-1189 1937-3333. DOI: 10.1109/RBME.2015.2390646. URL: https://www.ncbi.nlm.nih.gov/pubmed/25594979%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4813663/.

Srivastava, Devjit (2017). “Chronic post-amputation pain: peri-operative management - Review”. In: *British journal of pain* 11.4, pp. 192–202. ISSN: 2049-4637 2049-4645. DOI: 10.1177/2049463717736492. URL: https://pubmed.ncbi.nlm.nih.gov/29123664%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5661696/.

Stansbury, L. G. et al. (2008). “Amputations in U.S. military personnel in the current conflicts in Afghanistan and Iraq”. In: *J Orthop Trauma* 22.1, pp. 43–6. ISSN: 0890-5339 (Print) 0890-5339. DOI: 10.1097/BOT.0b013e31815b35aa.

Stepien, Jacqueline M. et al. (2007). “Activity Levels Among Lower-Limb Amputees: Self-Report Versus Step Activity Monitor”. In: 88.7, pp. 896–900. ISSN: 0003-9993. DOI: 10.1016/j.apmr.2007.03.016. URL: https://dx.doi.org/10.1016/j.apmr.2007.03.016.

Stolyarov, Roman, Gary Burnett, and Hugh Herr (2018). “Translational Motion Tracking of Leg Joints for Enhanced Prediction of Walking Tasks”. In: *IEEE Transactions on Biomedical Engineering* 65.4, pp. 763–769. DOI: 10.1109/TBME.2017.2718528.

Storm, Fabio A, Ben W Heller, and Claudia Mazzà (2015). “Step Detection and Activity Recognition Accuracy of Seven Physical Activity Monitors”. In: *PLOS ONE* 10.3, e0118723. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0118723. URL: http://10.0.5.91/journal.pone.0118723%20https://dx.doi.org/10.1371/journal.pone.0118723.

Studenski, Stephanie (2011). “Gait Speed and Survival in Older Adults”. In: *JAMA* 305.1, p. 50. ISSN: 0098-7484. DOI: 10.1001/jama.2010.1923. URL: https://dx.doi.org/10.1001/jama.2010.1923.

Su, Ben-Yue et al. (2019). “A CNN-Based Method for Intent Recognition Using Inertial Measurement Units and Intelligent Lower Limb Prosthesis”. In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 27.5, pp. 1032–1042. ISSN: 1534-4320. DOI: 10.1109/tnsre.2019.2909585. URL: http://10.0.4.85/tnsre.2019.2909585%20https://dx.doi.org/10.1109/tnsre.2019.2909585.

Su, Po-Fu et al. (2008). “Differences in Gait Characteristics Between Persons With Bilateral Transtibial Amputations, Due to Peripheral Vascular Disease and Trauma, and Able-Bodied Ambulators”. In: *Archives of Physical Medicine and Rehabilitation* 89.7, pp. 1386–1394. ISSN: 0003-9993. DOI: 10.1016/j.apmr.2007.10.050. URL: http://10.0.3.248/j.apmr.2007.10.050%20https://dx.doi.org/10.1016/j.apmr.2007.10.050.

Subasi, Abdulhamit et al. (2018). “Sensor Based Human Activity Recognition Using Adaboost Ensemble Classifier”. In: *Procedia Computer Science* 140, pp. 104–111. ISSN: 1877-0509. DOI: https://doi.org/10.1016/j.procs.2018.10.298. URL: http://www.sciencedirect.com/science/article/pii/S1877050918319719.

Subramaniam, Sophini et al. (2022). “Insole-Based Systems for Health Monitoring: Current Solutions and Research Challenges”. In: *Sensors* 22.2, p. 438. ISSN: 1424-8220. DOI: 10.3390/s22020438.

Sun, Jian et al. (2018). “Sequential Human Activity Recognition Based on Deep Convolutional Network and Extreme Learning Machine Using Wearable Sensors”. In: *Journal of Sensors* 2018, pp. 1–10. ISSN: 1687-725X. DOI: 10.1155/2018/8580959. URL: http://10.0.4.131/2018/8580959%20https://dx.doi.org/10.1155/2018/8580959.

Sun, Jie et al. (1996). “The influence of surface slope on human gait characteristics: a study of urban pedestrians walking on an inclined surface”. In: *Ergonomics* 39.4, pp. 677–692. ISSN: 0014-0139. DOI: 10.1080/00140139608964489. URL: http://10.0.4.56/00140139608964489%20https://dx.doi.org/10.1080/00140139608964489.

Suto, Jozsef, Stefan Oniga, and Petrica Pop Sitar (2016). "Feature Analysis to Human Activity Recognition". In: *International Journal of Computers Communications & Control* 12.1, p. 116. ISSN: 1841-9836. DOI: 10.15837/ijccc.2017.1.2787. URL: http://10.0.61.221/ijccc.2017.1.2787%20https://dx.doi.org/10.15837/ijccc.2017.1.2787.

Suto, Jozsef, Stefan Oniga, and Petrica Pop Sitar (2016). "Comparison of wrapper and filter feature selection algorithms on human activity recognition". In: *2016 6th International Conference on Computers Communications and Control (ICCCC)*. IEEE. DOI: 10.1109/icccc.2016.7496749. URL: http://10.0.4.85/icccc.2016.7496749%20https://dx.doi.org/10.1109/ICCCC.2016.7496749.

Performance of an activity monitor integrated into a microprocessor knee (2018) (Vancouver, Canada). Vol. 1. CPOJ. DOI: 10.33137/cpoj.v1i2.32031. URL: https://dx.doi.org/10.33137/cpoj.v1i2.32031.

Takeuchi, S et al. (2009). "Human Activity Recognition Based on Acceleration Information". In: *2009 IEICE Tech. Rep.* Vol. 108. 453, pp. 229–234. URL: https://www.ieice.org/ken/paper/20090303Oakt/eng/.

Tamura, T et al. (1997). "Classification of acceleration waveforms during walking by wavelet transform." In: *Methods of information in medicine* 36.4-5, pp. 356–359. ISSN: 0026-1270 (Print).

Tao, Dapeng et al. (2016). "Ensemble Manifold Rank Preserving for Acceleration-Based Human Activity Recognition". In: *IEEE Transactions on Neural Networks and Learning Systems* 27.6, pp. 1392–1404. ISSN: 2162-237X. DOI: 10.1109/tnnls.2014.2357794. URL: http://10.0.4.85/tnnls.2014.2357794%20https://dx.doi.org/10.1109/TNNLS.2014.2357794.

Taylor, Denise (2014). "Physical activity is medicine for older adults: Table 1". In: *Postgraduate Medical Journal* 90.1059, pp. 26–32. ISSN: 0032-5473. DOI: 10.1136/postgradmedj-2012-131366. URL: https://dx.doi.org/10.1136/postgradmedj-2012-131366.

TensorFlow (n.d.). Accessed: 2021-09. URL: https://www.tensorflow.org/.

Theeven, P. J. et al. (2012). "Influence of advanced prosthetic knee joints on perceived performance and everyday life activity level of low-functional persons with a transfemoral amputation or knee disarticulation". In: *J Rehabil Med* 44.5, pp. 454–61. ISSN: 1650-1977. DOI: 10.2340/16501977-0969.

Tibshirani, Robert, Guenther Walther, and Trevor Hastie (2001). "Estimating the number of clusters in a data set via the gap statistic". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63.2, pp. 411–423. ISSN: 1369-7412. DOI: https://doi.org/10.1111/1467-9868.00293. URL: https://doi.org/10.1111/1467-9868.00293.

Trabelsi, Dorra et al. (2013). "An Unsupervised Approach for Automatic Activity Recognition Based on Hidden Markov Model Regression". In: *IEEE Transactions on Automation*

Science and Engineering 10.3, pp. 829–835. ISSN: 1545-5955. DOI: 10.1109/tase.2013.2256349. URL: http://10.0.4.85/tase.2013.2256349%20https://dx.doi.org/10.1109/tase.2013.2256349.

Tudor-Locke, Catrine et al. (2020). "Walking cadence (steps/min) and intensity in 41 to 60-year-old adults: the CADENCE-adults study". In: *International Journal of Behavioral Nutrition and Physical Activity* 17.1. ISSN: 1479-5868. DOI: 10.1186/s12966-020-01045-z. URL: http://10.0.4.162/s12966-020-01045-z%20https://dx.doi.org/10.1186/s12966-020-01045-z.

Twomey, Niall et al. (2018). *A Comprehensive Study of Activity Recognition Using Accelerometers*. DOI: 10.3390/informatics5020027.

Vageskar, Eirik (2017). "Activity Recognition for Stroke Patients". MA thesis. Norweigan University of Science and Technology.

Vaizman, Y, K Ellis, and G Lanckriet (2017). "Recognizing Detailed Human Context in the Wild from Smartphones and Smartwatches". In: *IEEE Pervasive Computing* 16.4, pp. 62–74. ISSN: 1558-2590. DOI: 10.1109/MPRV.2017.3971131.

Van Den Akker, Lizanne E et al. (2020). "Determinants of physical activity in wheelchair users with spinal cord injury or lower limb amputation: perspectives of rehabilitation professionals and wheelchair users". In: *Disability and Rehabilitation* 42.14, pp. 1934–1941. ISSN: 0963-8288. DOI: 10.1080/09638288.2019.1577503. URL: http://10.0.4.56/09638288.2019.1577503%20https://dx.doi.org/10.1080/09638288.2019.1577503.

Van Den Berg-Emons, Rita J. et al. (2011). "Validation of the Physical Activity Scale for Individuals With Physical Disabilities". In: 92.6, pp. 923–928. ISSN: 0003-9993. DOI: 10.1016/j.apmr.2010.12.006. URL: https://dx.doi.org/10.1016/j.apmr.2010.12.006.

Van Hees, Vincent T et al. (2013). "Separating Movement and Gravity Components in an Acceleration Signal and Implications for the Assessment of Human Daily Physical Activity". In: *PLoS ONE* 8.4, e61691. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0061691. URL: http://10.0.5.91/journal.pone.0061691%20https://dx.doi.org/10.1371/journal.pone.0061691.

Vanicek, Natalie et al. (2009). "Gait patterns in transtibial amputee fallers vs. non-fallers: Biomechanical differences during level walking". In: *Gait & Posture* 29.3, pp. 415–420. ISSN: 0966-6362. DOI: 10.1016/j.gaitpost.2008.10.062. URL: http://10.0.3.248/j.gaitpost.2008.10.062%20https://dx.doi.org/10.1016/j.gaitpost.2008.10.062.

Voloshina, Alexandra S et al. (2013). "Biomechanics and energetics of walking on uneven terrain". In: *Journal of Experimental Biology* 216.21, pp. 3963–3970. ISSN: 1477-9145. DOI: 10.1242/jeb.081711. URL: http://10.0.4.218/jeb.081711%20https://dx.doi.org/10.1242/jeb.081711.

Wall, R., P. Novotny-Joseph, and T. E. Macnamara (1985). "Does preamputation pain influence phantom limb pain in cancer patients?" In: *South Med J* 78.1, pp. 34–6. ISSN: 0038-4348 (Print) 0038-4348. DOI: 10.1097/00007611-198501000-00009.

Walse, Kishor H., Rajiv V. Dharaskar, and Vilas M. Thakare (2016). "PCA Based Optimal ANN Classifiers for Human Activity Recognition Using Mobile Sensors Data". In: *Proceedings of First International Conference on Information and Communication Technology for Intelligent Systems: Volume 1*. Ed. by Suresh Chandra Satapathy and Swagatam Das. Springer International Publishing, pp. 429–436. ISBN: 978-3-319-30933-0.

Wang, Ning et al. (2007). "Accelerometry Based Classification of Walking Patterns Using Time-frequency Analysis". In: *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE. DOI: 10.1109/iembs.2007.4353438. URL: http://10.0.4.85/iembs.2007.4353438%20https://dx.doi.org/10.1109/iembs.2007.4353438.

Wang, Shuangquan et al. (2005). "Human Activity Recognition with User-Free Accelerometers in the Sensor Networks". In: *2005 International Conference on Neural Networks and Brain*. IEEE. DOI: 10.1109/icnnb.2005.1614831. URL: http://10.0.4.85/icnnb.2005.1614831%20https://dx.doi.org/10.1109/ICNNB.2005.1614831.

Wang, Yan, Shuang Cang, and Hongnian Yu (2016). "A review of sensor selection, sensor devices and sensor deployment for wearable sensor-based human activity recognition systems". In: *2016 10th International Conference on Software, Knowledge, Information Management & Applications (SKIMA)*. IEEE. DOI: 10.1109/skima.2016.7916228. URL: http://10.0.4.85/skima.2016.7916228%20https://dx.doi.org/10.1109/SKIMA.2016.7916228.

Warburton, D. E. R. (2006). "Health benefits of physical activity: the evidence". In: *Canadian Medical Association Journal* 174.6, pp. 801–809. ISSN: 0820-3946. DOI: 10.1503/cmaj.051351. URL: https://dx.doi.org/10.1503/cmaj.051351.

Washburn, R. A. et al. (2002). "The physical activity scale for individuals with physical disabilities: development and evaluation". In: *Arch Phys Med Rehabil* 83.2, pp. 193–200. ISSN: 0003-9993 (Print) 0003-9993.

Watters, Kirsty and Sarah Deans (2015). *Physical activity perceptions of prosthesis users : an interpretative phenomenological analysis*. FRA. URL: https://strathprints.strath.ac.uk/52687/.

Wei, Yu et al. (2018). "Sensor selection scheme in activity recognition based on hierarchical feature reduction". In: *International Journal of Distributed Sensor Networks* 14.8, p. 1550147718793801. DOI: 10.1177/1550147718793801.

Wetterhahn, Kristin, Carolyn Hanson, and Charles Levy (2002). "Effect of Participation in Physical Activity on Body Image of Amputees". In: *American journal of physical* *medicine & rehabilitation / Association of Academic Physiatrists* 81, pp. 194–201. DOI: 10.1097/00002060-200203000-00007.

Wilcoxon, Frank (1945). “Individual Comparisons by Ranking Methods”. In: *Biometrics Bulletin* 1.6, pp. 80–83. ISSN: 00994987. URL: http://www.jstor.org/stable/3001968.

Winkler, Elisabeth A H et al. (2016). “Identifying adults’ valid waking wear time by automated estimation in activPAL data collected with a 24 h wear protocol”. In: *Physiological Measurement* 37.10, pp. 1653–1668. ISSN: 0967-3334. DOI: 10.1088/0967-3334/37/10/1653. URL: http://10.0.4.64/0967-3334/37/10/1653%20https://dx.doi.org/10.1088/0967-3334/37/10/1653.

Wolf, Erik J et al. (2012). “Assessment of transfemoral amputees using C-Leg and Power Knee for ascending and descending inclines and steps”. In: *The Journal of Rehabilitation Research and Development* 49.6, p. 831. ISSN: 0748-7711. DOI: 10.1682/jrrd.2010.12.0234. URL: http://10.0.6.146/jrrd.2010.12.0234%20https://dx.doi.org/10.1682/jrrd.2010.12.0234.

World Health Organization (n.d.). *Physical Activity*. URL: https://www.who.int/dietphysicalactivity/pa/en/.

Woznowski, Przemyslaw et al. (2016). *A Human Activity Recognition Framework for Healthcare Applications: Ontology, Labelling Strategies, and Best Practice*, pp. 369–377. DOI: 10.5220/0005932503690377.

Wu, Hsin-I et al. (2000). “Modelling animal movement as a persistent random walk in two dimensions: expected magnitude of net displacement”. In: 132.1-2, pp. 115–124. ISSN: 0304-3800. DOI: 10.1016/s0304-3800(00)00309-4. URL: http://10.0.3.248/s0304-3800(00)00309-4%20https://dx.doi.org/10.1016/S0304-3800(00)00309-4.

Wu, Jin-Ming et al. (2019). “Wearable-Based Mobile Health App in Gastric Cancer Patients for Postoperative Physical Activity Monitoring: Focus Group Study”. In: *JMIR mHealth and uHealth* 7.4, e11989. ISSN: 2291-5222. DOI: 10.2196/11989. URL: http://10.0.8.148/11989%20https://dx.doi.org/10.2196/11989.

Wu, Yuanxin et al. (2018). “Dynamic Magnetometer Calibration and Alignment to Inertial Sensors by Kalman Filtering”. In: *IEEE Transactions on Control Systems Technology* 26.2, pp. 716–723. ISSN: 1063-6536. DOI: 10.1109/tcst.2017.2670527. URL: https://dx.doi.org/10.1109/TCST.2017.2670527.

Xia, Kun et al. (2020). “Racquet Sports Recognition Using a Hybrid Clustering Model Learned from Integrated Wearable Sensor”. In: *Sensors* 20.6, p. 1638. ISSN: 1424-8220. DOI: 10.3390/s20061638. URL: http://10.0.13.62/s20061638%20https://dx.doi.org/10.3390/s20061638.

Xu, Lu et al. (2017). “Human activity recognition based on random forests”. In: *2017 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Dis-*

covery (ICNC-FSKD). IEEE. DOI: 10.1109/fskd.2017.8393329. URL: http://10.0.4.85/fskd.2017.8393329%20https://dx.doi.org/10.1109/FSKD.2017.8393329.

Yamada, M. et al. (2012). "Pedometer-Based Behavioral Change Program Can Improve Dependency in Sedentary Older Adults: A Randomized Controlled Trial". In: *J Frailty Aging* 1.1, pp. 39–44. ISSN: 2260-1341 (Print) 2260-1341. DOI: 10.14283/jfa.2012.7.

Yang, Jhun-Ying et al. (2007). "Activity Recognition Using One Triaxial Accelerometer: A Neuro-fuzzy Classifier with Feature Reduction". In: vol. 4740, pp. 395–400. ISBN: 978-3-540-74872-4. DOI: 10.1007/978-3-540-74873-1_47.

Yang, Shuozhi and Qingguo Li (2012). "Inertial Sensor-Based Methods in Walking Speed Estimation: A Systematic Review". In: *Sensors* 12.5, pp. 6102–6116. ISSN: 1424-8220. DOI: 10.3390/s120506102. URL: http://10.0.13.62/s120506102%20https://dx.doi.org/10.3390/s120506102.

Yeager, Richard A. et al. (1995). "Deep vein thrombosis associated with lower extremity amputation". In: *Journal of Vascular Surgery* 22.5, pp. 612–615. ISSN: 0741-5214. DOI: 10.1016/s0741-5214(95)70048-x. URL: https://dx.doi.org/10.1016/s0741-5214(95)70048-x.

Yi Wang, Xiao and Xiu Yun Meng (2017). "Research on Time-series Modeling and Filtering Methods for MEMS Gyroscope Random Drift Error". In: *IOP Conference Series: Materials Science and Engineering* 187, p. 012005. ISSN: 1757-8981 1757-899X. DOI: 10.1088/1757-899x/187/1/012005. URL: http://dx.doi.org/10.1088/1757-899X/187/1/012005.

Yin, Jie, Qiang Yang, and J J Pan (2008). "Sensor-Based Abnormal Human-Activity Detection". In: *IEEE Transactions on Knowledge and Data Engineering* 20.8, pp. 1082–1090. ISSN: 1041-4347. DOI: 10.1109/tkde.2007.1042. URL: http://10.0.4.85/tkde.2007.1042%20https://dx.doi.org/10.1109/TKDE.2007.1042.

Yorston, Lisa C., Gregory S. Kolt, and Richard R. Rosenkranz (2012). "Physical Activity and Physical Function in Older Adults: The 45 and Up Study". In: *Journal of the American Geriatrics Society* 60.4, pp. 719–725. ISSN: 0002-8614. DOI: 10.1111/j.1532-5415.2012.03906.x. URL: https://dx.doi.org/10.1111/j.1532-5415.2012.03906.x.

Young, J et al. (2015). *Measuring Change: an Introduction to Clinical Outcome Measures in Prosthetics and Orthotics*. URL: https://www.bapo.com/wp-content/uploads/2019/02/Measuring-Change-BAPO-website.pdf.

Young, Joshua, Lynne Rowley, and Simon Lalor (2018). *Use of Outcome Measures Among Prosthetists and Orthotists in the United Kingdom*. Vol. 30. DOI: 10.1097/JPO.0000000000000198.

Yuan, Chunhui and Haitao Yang (2019). *Research on K-Value Selection Method of K-Means Clustering Algorithm*. DOI: 10.3390/j2020016.

Zappi, Piero et al. (2007). “Activity recognition from on-body sensors by classifier fusion: sensor scalability and robustness”. In: *2007 3rd International Conference on Intelligent Sensors, Sensor Networks and Information*, pp. 281–286. DOI: 10.1109/ISSNIP.2007.4496857.

Zebin, Tahmina et al. (2018). “Human activity recognition from inertial sensor time-series using batch normalized deep LSTM recurrent networks”. In: *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE. DOI: 10.1109/embc.2018.8513115. URL: http://10.0.4.85/embc.2018.8513115%20https://dx.doi.org/10.1109/embc.2018.8513115.

Zhang, Jianhua, Chen Ling, and Sunan Li (2019). “EMG Signals based Human Action Recognition via Deep Belief Networks”. In: *IFAC-PapersOnLine* 52.19, pp. 271–276. ISSN: 2405-8963. DOI: https://doi.org/10.1016/j.ifacol.2019.12.108. URL: http://www.sciencedirect.com/science/article/pii/S2405896319319433.

Zhang, Kuangen et al. (2019a). “Environmental Features Recognition for Lower Limb Prostheses Toward Predictive Walking”. In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 27.3, pp. 465–476. DOI: 10.1109/TNSRE.2019.2895221.

Zhang, Kuangen et al. (2019b). “Sequential Decision Fusion for Environmental Classification in Assistive Walking”. In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 27.9, pp. 1780–1790. DOI: 10.1109/TNSRE.2019.2935765.

Zhang, Mi and Alexander A. Sawchuk (2011). “A Feature Selection-Based Framework for Human Activity Recognition Using Wearable Multimodal Sensors”. In: *Proceedings of the 6th International Conference on Body Area Networks*. BodyNets ’11. Beijing, China: ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), pp. 92–98. ISBN: 9781936968299.

Zhao, Kunlun et al. (2013). “Healthy: A Diary System Based on Activity Recognition Using Smartphone”. In: *2013 IEEE 10th International Conference on Mobile Ad-Hoc and Sensor Systems*. IEEE. DOI: 10.1109/mass.2013.14. URL: http://10.0.4.85/mass.2013.14%20https://dx.doi.org/10.1109/mass.2013.14.

Zhao, Yu et al. (2017). “Deep Residual Bidir-LSTM for Human Activity Recognition Using Wearable Sensors”. In: *arXiv pre-print server*. DOI: None. URL: arxiv:1708.08989%20https://arxiv.org/abs/1708.08989.

Zhao, Zhenyu, Radhika Anand, and Mallory Wang (2019). “Maximum Relevance and Minimum Redundancy Feature Selection Methods for a Marketing Machine Learning Platform”. In: *2019 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*. IEEE. DOI: 10.1109/dsaa.2019.00059. URL: http://10.0.4.85/dsaa.2019.00059%20https://dx.doi.org/10.1109/dsaa.2019.00059.

Zheng, Yuhuang (2015). “Human Activity Recognition Based on the Hierarchical Feature Selection and Classification Framework”. In: *Journal of Electrical and Computer Engi-*

neering 2015. Ed. by Sos Agaian, p. 140820. ISSN: 2090-0147. DOI: 10.1155/2015/140820. URL: https://doi.org/10.1155/2015/140820.

Zhu, R. and Z. Zhou (2004). "A Real-Time Articulated Human Motion Tracking Using Tri-Axis Inertial/Magnetic Sensors Package". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 12.2, pp. 295–302. ISSN: 1534-4320. DOI: 10.1109/tnsre.2004.827825. URL: https://dx.doi.org/10.1109/TNSRE.2004.827825.

Ziegler-Graham, Kathryn et al. (2008). "Estimating the Prevalence of Limb Loss in the United States: 2005 to 2050". In: *Archives of Physical Medicine and Rehabilitation* 89.3, pp. 422–429. ISSN: 0003-9993. DOI: 10.1016/j.apmr.2007.11.005. URL: https://dx.doi.org/10.1016/j.apmr.2007.11.005.

Zijlstra, Wiebren (2004). "Assessment of spatio-temporal parameters during unconstrained walking." In: *European journal of applied physiology* 92.1-2, pp. 39–44. ISSN: 1439-6319 (Print). DOI: 10.1007/s00421-004-1041-5.

APPENDIX

Appendix A

Interview and Focus Group Formats for Individuals with Lower Limb Amputation and Healthcare Professionals

A.1 Overview

This appendix contains the original text document for the structure of interviews and focus group conducted in chapter 3 for HCPs and ILLAs. The questions presented in this document are only a rough representation of the questions actually used in the sessions; there were numerous unscripted follow-up questions used in each session. The theming of the questions used in the interviews and focus group of the ILLAs were identical, and have thus been represented with a single set of questions.

Interview Format- Healthcare Professionals

Section 1 – Job Details

- What is/was your job title?
- How long have/had you held this position (or another position similar to your current one)?
- Do you work or have previously worked with lower limb loss patients? What kinds of lower limb patients do you generally work with most? (transtibial vs transfemoral, unilateral vs. bilateral), any unique cases? (e.g one leg is transtibial, other is transfemoral)
- How are your rehabilitation sessions structured? Do you interact with patients in a group or individually?
- How frequently do/did you see each patient?
- How long is/was each session? Do you feel the current length is adequate?

Section 2 – Assessment of Patient Activity

- Do you use or have you considered using physical activity assessment with your patients? If not why not?
- [Skip if answer to above is no] When assessing a patient in your clinic what are the key observations and metrics do you use?
- If no external monitoring (self-report, measuring devices) is mentioned, why do you not do use this?

Section 3 – Desires for activity monitoring outcomes

- What kinds of physical activities would you expect your patients to carry out?
- Are there certain types of physically activity in which the patient would, on a regular basis, remove their prosthetic before carrying out the exercise?
- Can you name any forms of physical activity or exercise that you desire to be measured via a sensor (e.g pedometer, Fitbit etc.)?

Section 4 – Communication with Clients

- How do you communicate your observations to patients in regard to their gait quality or fitness levels?
- What language do you use (simple language, use of medical terminology)?
- Do you use video or diagrams?
- Do you prescribe exercise or training-based goals?
- [if prescribe exercise] What methods do you use to check they're doing these exercises often enough and correctly?
- Do you experience any difficulties communicating with your patients? (e.g do you feel that your patients lie or are unreliable about reporting their activities, or do you feel like the amputee struggles to understand what you want them to do?)

Section 5 – Desires for a clinical activity monitoring system

- In this system, which do you think would be the best method to communicate to participants about what PA they should be carrying out?
- Would you think a combination of formats could be beneficial?

- How frequently do you think the system should communicate physical activity progress to your patients?
- Theoretically, to what extent would you like to have manual input with the system? (i.e would you want to manually interact or use automated responses)
- Are there any difficulties you foresee with encouraging the patient to carry out the activity while wearing the leg prosthesis?
- Do you have any concerns about the privacy the patient may have with this system?

Interview/Focus Group Format – Lower Limb Amputees

Section 1 – Amputation, Rehabilitation Experiences

- Could you describe the technical details of your amputation (e.g unilateral/bilateral, transtibial/transfemoral)
- When were you first fitted for your prosthesis?
- (if comfortable sharing) what was the cause of your lower limb amputation?
- What is your current profession?
    - Did your amputation force you to change careers?
- How many physiotherapy sessions have you had? (if any)
- How long since your first physiotherapy session?
- How frequently do you have a physiotherapy session on average?
- How long was each session?
- Do you have sessions individually (with a HCP) or as a group?
- What were the most helpful parts of your sessions?
- What were the least helpful parts?
- What would have made the sessions better?
- During your rehabilitation what were your main concerns?

Section 2 – Interactions with healthcare professionals and health assessments

- When meeting with your physiotherapists, in what ways are you assessed? (example 5 minute walk test)
- How is information about your activity given to you? (Verbal, digital presentation, hard copies etc.)
- Depth of feedback
    - Are recommendations specific or general? (e.g do more exercise vs. do x amount of certain exercise)
    - Do HCPs set out (or did set out if no longer in rehabilitation) daily, weekly or monthly goals for things like exercises or step count?
    - Do you sometimes struggle to understand how you need to meet their requirements?
    - On average, do you know (or strongly believe) if you are meeting the requirements of physical activity set out by your HCP?
- On average, do you feel that the assessment given by the HCP overestimates, underestimates, or accurately reflects how active you think you actually are?
- Influence of feedback on goals
    - When you receive good feedback from your HCP, do you feel that makes you more likely to keep your target for the next month, less likely, or has no effect?
    - The same question as above but for negative feedback / failing to meet standards

Section 3 – Physical Activity

- Are you aware of your K-level?
    - (if not give description of each K-level and ask which sounds the most appropriate)
- How frequently do you walk (if at all)?
- Where do you go on your walks?
- What other physical activities do you perform?
- How frequently do you perform these physical activities?
- Are there any other physical activities you would like to carry out? If so, do you feel your amputation impairs you from doing so?
- How frequently do you wear your prosthesis when you perform these exercises?
- [from section 2, if someone said they had physical activity assessment] how does your physiotherapist assess whether you've carried out this activity
- Are there any other physical activities I've not mentioned that you carry out?
    - How frequently do you perform those physical activities?
    - How frequently do you wear your prosthesis when you perform these exercises?
    - Does wearing the prosthesis make it easier, harder or the same to do than without
- Which of these activities discussed do you think are the easiest and the hardest to carry out on a daily basis?

Section 4 – Experience with Technology

*[1-5] on a scale of 1-5, with 1 being not comfortable at all, and 5 meaning very comfortable.

- Computers
    - Do you own a computer?
    - [1-5]* How comfortable are you with checking your e-mails?
    - [1-5] How comfortable are you with using the internet?
    - [If no] Would you ever consider using a computer in the future?
- Smartphones/Tablets
    - Do you own a smartphone or a tablet? (Alternatively, just a mobile phone)
    - [1-5] How comfortable are you with reading text messages?
    - [1-5] How comfortable are you with checking your e-mails?
    - [1-5] How comfortable are you with using the internet?
    - [1-5] How comfortable are you with downloading and using apps?
    - [If no] Would you ever consider using one in the future?
- Smart Watch
    - Have you ever used a smart watch?
    - [If yes, 1-5] Are you comfortable operating it?
    - [If no] Would you ever consider using one?
- Fitness Tracker
    - Have you ever used a fitness tracker (e.g Fitbit)?
    - [If yes, 1-5] Are you comfortable operating it?
    - [If no] Would you ever consider using one?

- If you had to monitor and track your fitness, Of the technology discussed, which device would you prefer for this purpose?

Appendix B

Feature Extraction: Description of Time-Consuming Functions

This appendix explains the functions of Table 5.5 in Chapter 5 in more detail. The explanations of the functions are broken into two categories; in-built MATLAB functions and functions created by the author.

Inbuilt Matlab functions:

- **wavedec:** Returns the wavelet decomposition of a one-dimensional signal. Is used in the calculation of all wavelet domain features.
- **dwt:** Performs a discrete wavelet transform on a one-dimensional signal, is called automatically when running *"wavedec"*
- **wfilters:** Creates four lowpass and highpass, decomposition and reconstruction filters associated with a defined wavelet structure, is called automatically when running *"wavedec"*
- **prctile:** Returns the 25th and 75th percentile of the window segment
- **stft:**Returns the Short-Time Fourier transform of a one-dimensional signal, is used in the calculation of Mel-Frequency Cepstral Coefficients (MFCCs), see *"cepstral_feature_function"*
- **wavemngr:** Adds or removes wavelets, is called when running *"wfilters"*

- **stftParser:** No documentation appears to exist for this function, presumably a low-level operation that gets called when *"stft"* is called.

- **appcoef:** Extracts level 1 approximation coefficients from a decomposed wavelet signal. Is used in the calculation of wavelet domain features.

Created functions:

- **cepstral_feature_function:** Calculates the first 13 cepstral coefficients of a one-dimensional signal, each of which are used as a separate feature. A short-term Fourier transform is applied to the window segment, after which the absolute values of the transformed signal are taken. Matlab extracts the cepstral coefficients using an in-built function *"cepstralCoefficients"*. By adding the calculation times for "*stft*" and "*stftParser*", it is evident that the Short Time Fourier Transform is responsible for consuming the most time within this function.

- **AutoCorr_Features:** Extracts the first seasonal lag and the squared sum of the first 10 autocorrelation coefficients. Coefficients are extracted by using the Matlab autocorrelation function *"autocorr"* (this is the function that is causing the time consumption), then the resulting features can be extracted through simple vector and arithmetic operations

Appendix C

Supervised Learning of Human Activity Recognition for Non-amputated Individuals and Individuals with Lower Limb Amputation in Free-Living Conditions - Additional Figures

This appendice contains the confusion matrices from Experiment #4 of Chapter 6.3.4. Confusion charts for both SVM and LSTM models are presented for each individual, and through each training method. All experiments are conducted at level 1 label resolution.

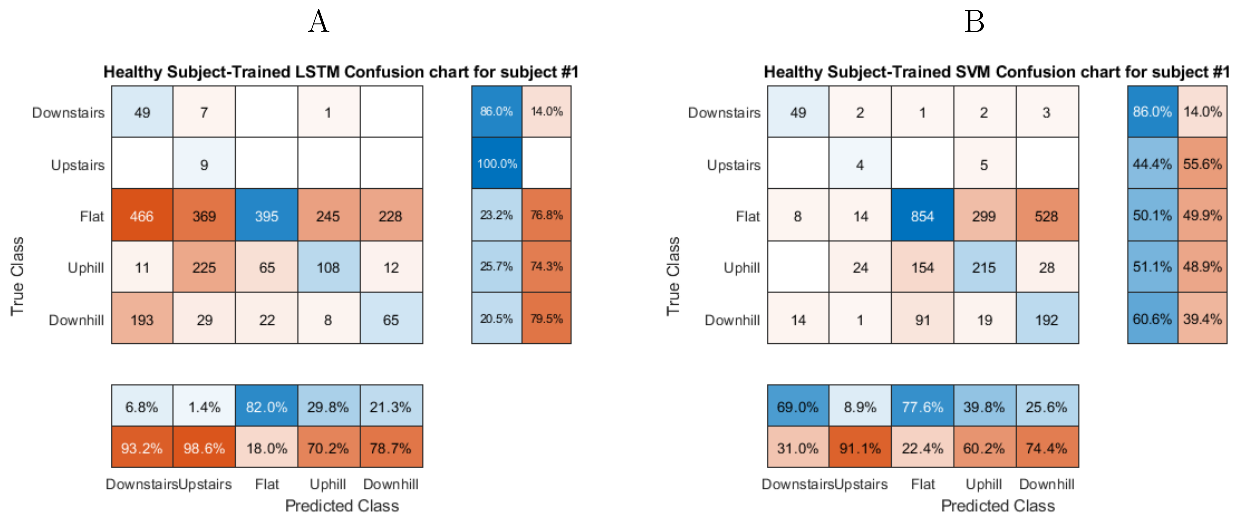


Figure C.1 A: Confusion chart of LSTM classifier trained exclusively on non-amputated participant data, tested on participant A1 || B: Confusion chart of SVM classifier trained exclusively on non-amputated participant data, tested on participant A1

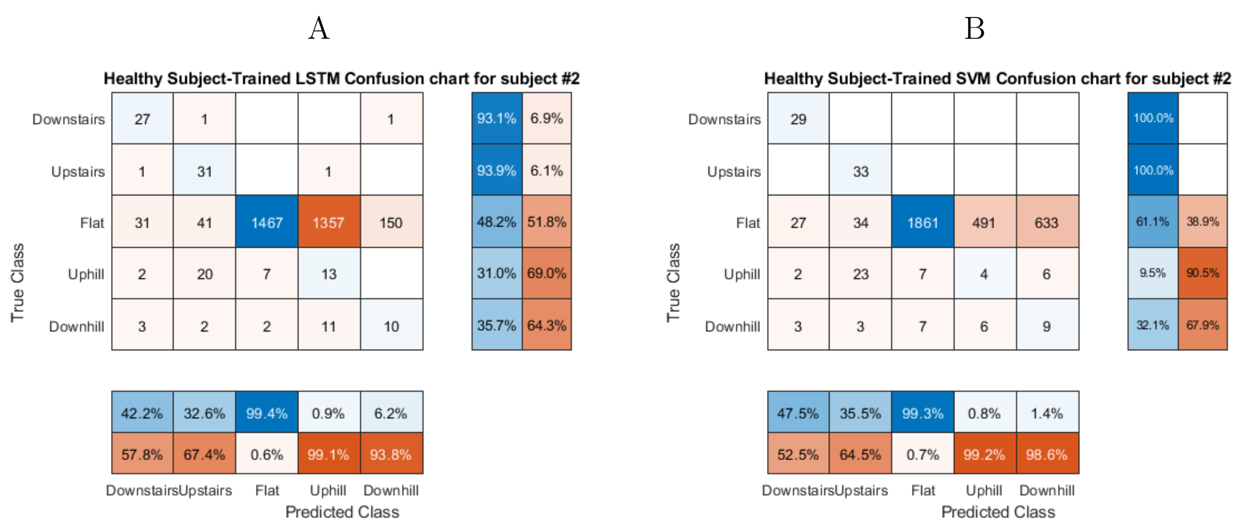


Figure C.2 A: Confusion chart of LSTM classifier trained exclusively on non-amputated participant data, tested on participant A2 || B: Confusion chart of SVM classifier trained exclusively on non-amputated participant data, tested on participant A2

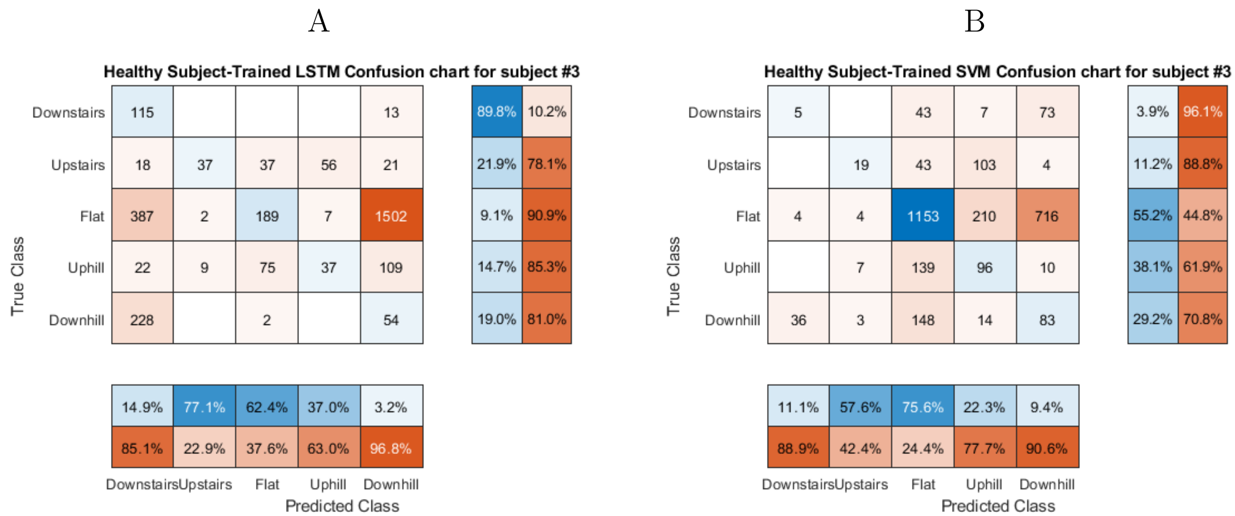


Figure C.3 A: Confusion chart of LSTM classifier trained exclusively on non-amputated participant data, tested on participant A3 || B: Confusion chart of SVM classifier trained exclusively on non-amputated participant data, tested on participant A3

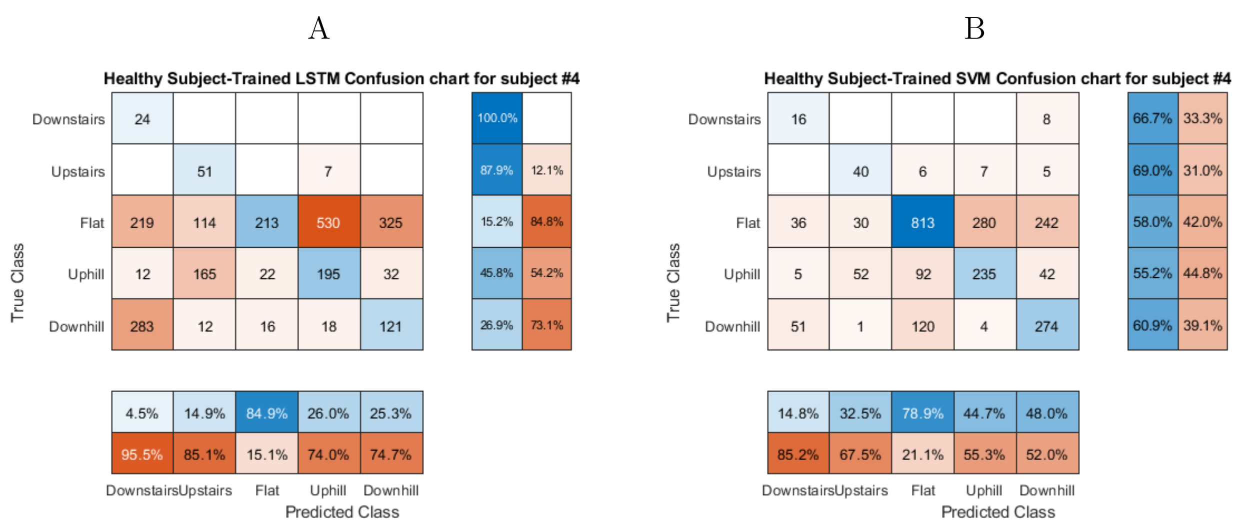


Figure C.4 A: Confusion chart of LSTM classifier trained exclusively on non-amputated participant data, tested on participant A4 || B: Confusion chart of SVM classifier trained exclusively on non-amputated participant data, tested on participant A4

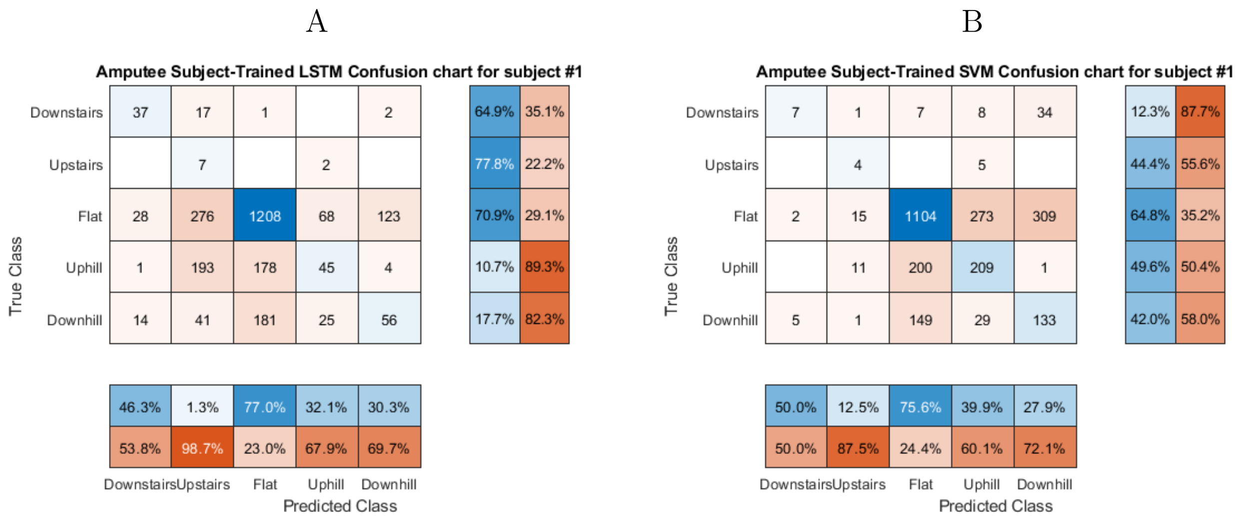


Figure C.5 A: Confusion chart of LSTM classifier trained exclusively on ILLA participant data, tested on participant A1 || B: Confusion chart of SVM classifier trained exclusively on ILLA participant data, tested on participant A1

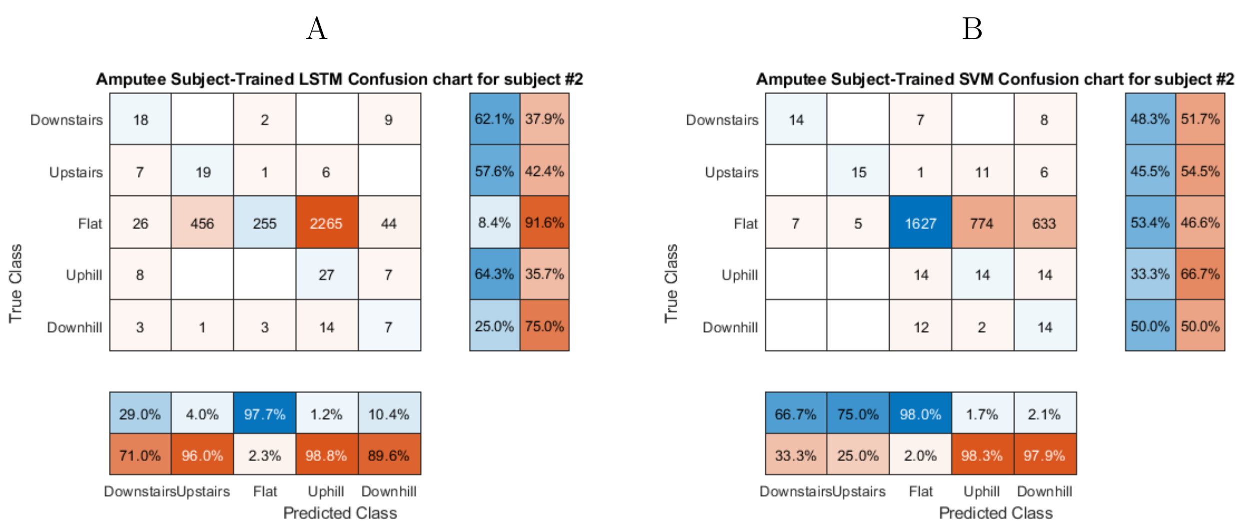


Figure C.6 A: Confusion chart of LSTM classifier trained exclusively on ILLA participant data, tested on participant A2 || B: Confusion chart of SVM classifier trained exclusively on ILLA participant data, tested on participant A2

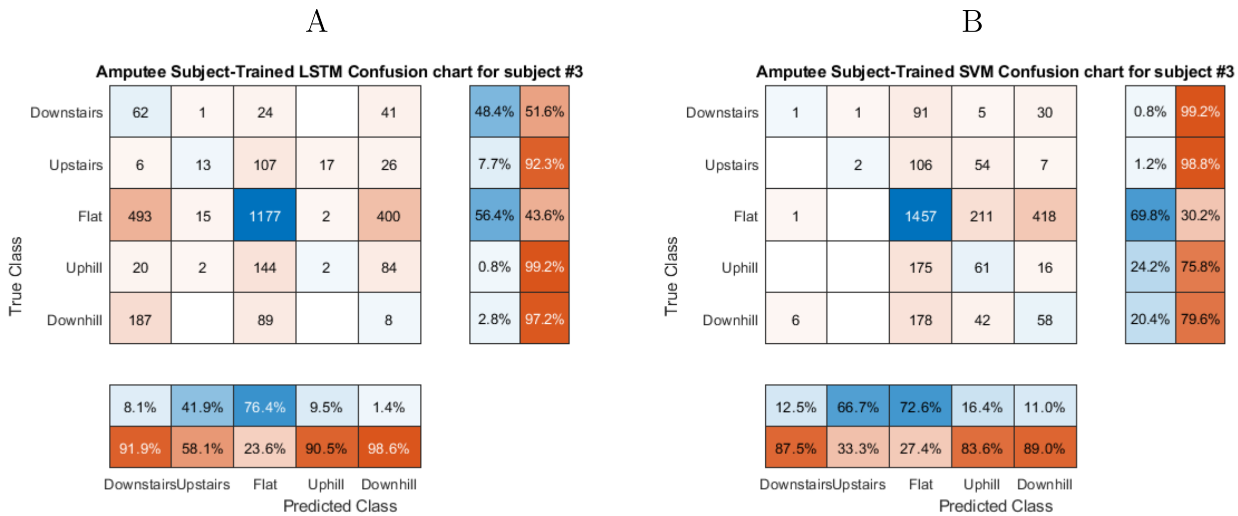


Figure C.7 A: Confusion chart of LSTM classifier trained exclusively on ILLA participant data, tested on participant A3 || B: Confusion chart of SVM classifier trained exclusively on ILLA participant data, tested on participant A3

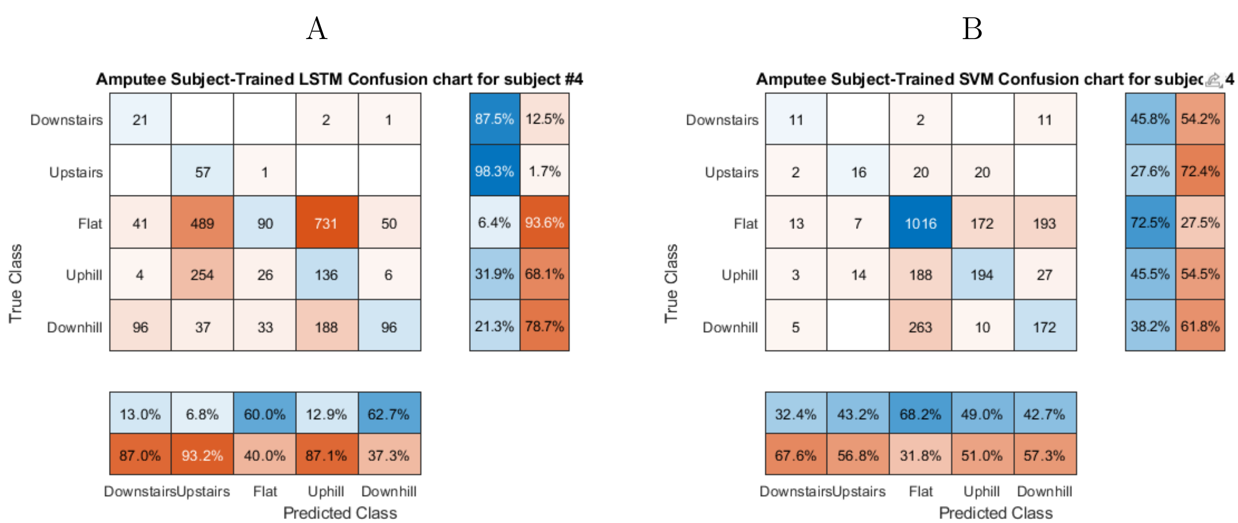


Figure C.8 A: Confusion chart of LSTM classifier trained exclusively on ILLA participant data, tested on participant A4 || B: Confusion chart of SVM classifier trained exclusively on ILLA participant data, tested on participant A4

Appendix D

Stair Detection Algorithm Pseudo Code

```
%Stair Clustering Algorithm Psuedo Code

for i = 1: numOfSubjects
    %Perform dimensionality reduction:
    tsne_model = tsne(subject_data(i));
    %Perform DBSCAN:
    DBSCAN_idx = dbscan(tsne_model,epsilon,minpts);
    %Remove outlier points from the model:
    tsne_model(DBSCAN_idx == -1) = [];
    DBSCAN_idx(DBSCAN_idx == -1) = [];
    %Algorithm has high chance of failure if dataset is too small, or there
    %are too many clusters. It also fails automatically if there is only 1
    %cluster identified (discounting outliers)
    cluster_elements = unique(DBSCAN_idx);
    if size(subject_data) < threshold || numel(cluster_elements) == 1 || numel(cluster_elements) > 10
        print("Error!")
        continue
    end
    %Locate the largest cluster and assign it as the walking cluster:
    walk_cluster_ID = max(tabulate(DBSCAN_idx));
    walk_idx = find(DBSCAN_idx == walk_cluster_ID);
    %Fit a 1-component Gaussian Mixture Model to the walking cluster:
    GMModel = fitgmdist(tsne_model(walk_idx),1);
    %For all stair candidate (stairCan) clusters, calculate the centroid and compactness:
    k = 1; %dynamic indice for stair candidate clusters
    for j = 1:numel(unique(DBSCAN_idx))
        if j ~= walk_cluster_ID
            stairCan_cluster_idx(k) = find(DBSCAN_idx == cluster_elements(j));
            stairCan_cluster_data(k) = tsne_model(stairCan_cluster_idx(k),:);
            stairCan_cluster_centroid(k) = centroid(stairCan_cluster_data(k));
            stairCan_cluster_compactness(k) = median(pdist(stairCan_cluster_data(k)).^-1);
            k = k + 1;
        else
            continue
        end
    end
    %For each of the stair candidate clusters, rank each cluster in order
    %of increasing compactness:
    [~,rank_idx] = sort(stairCan_cluster_compactness, 'ascend');
    %Assign a ranking factor, iota, to each cluster and calculate the
    %negative log likelihood of the cluster belonging to the central walk
    %cluster:
    for L = 1:k
        rank_factor = rank_idx(L);
        iota(L) = (rank_factor/rank_factor) + iota_scale*(rank_factor-1);
        [~,logLike(L)] = posterior(GMModel, stairCan_cluster_centroid(k));
    end
    %Calculate the probability that the stair candidate cluster is the
    %stair cluster:
    for L = 1:k
       stair_probability(L) = 1 - (1/(logLike(L) * iota(L)));
    end
    %The stair cluster is thus the stair cluster candidate with the highest
    %probability:
    [~,stair_cluster_ID] = max(stair_probability,1);
end
```

Figure D.1 Matlab pseudo code for stair detection described in Chapter 7.3.2.

Appendix E

Unsupervised Cluster Analysis of Walking Activity Data for Non-amputated Individuals and Individuals with Lower Limb Amputation - Additional Figures

This appendix demonstrates exemplary results of how activity labels were clustered by the tSNE algorithm (the ground truth) in Chapter 7, and the resulting cluster model developed by the proposed algorithm. This is applied for all subjects barring subject H1. Note that these figures are not an accurate representation of the numerical results obtained in this investigation (i.e the obtained NMIs of Table 7.8) and are simply for illustration purposes.

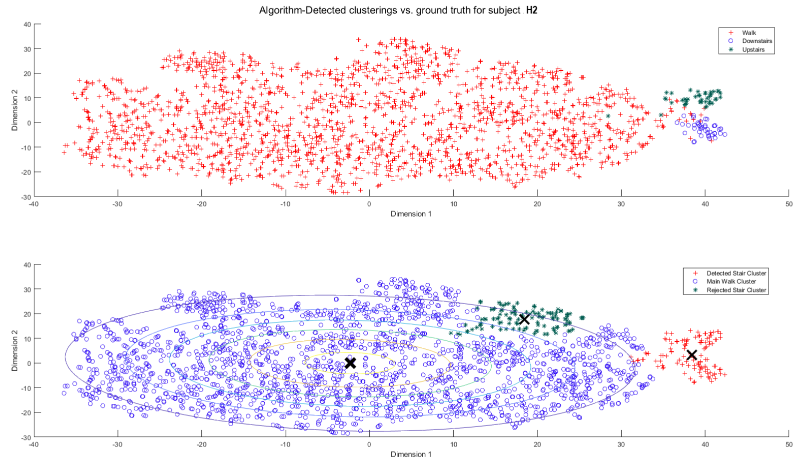


Figure E.1 Top: tSNE cluster model of participant H2, labelled with ground truth activity labels at label resolution 0. || Bottom: tSNE cluster model of participant H2, with activities identified by the proposed clustering algorithm.

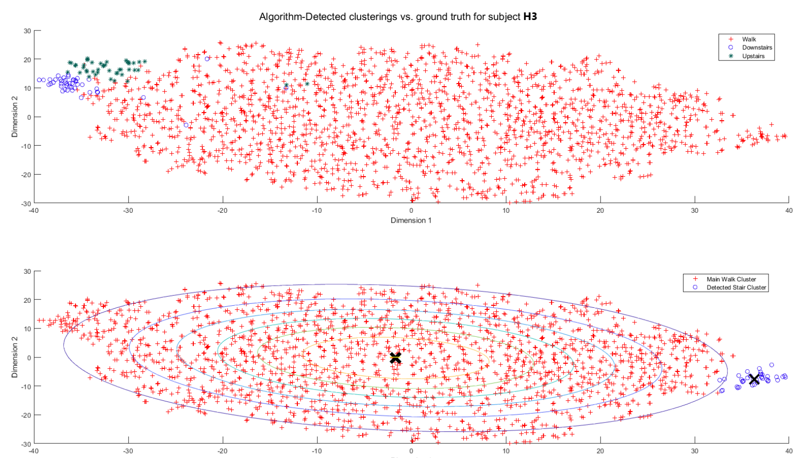


Figure E.2 Top: tSNE cluster model of participant H3, labelled with ground truth activity labels at label resolution 0. || Bottom: tSNE cluster model of participant H3, with activities identified by the proposed clustering algorithm.

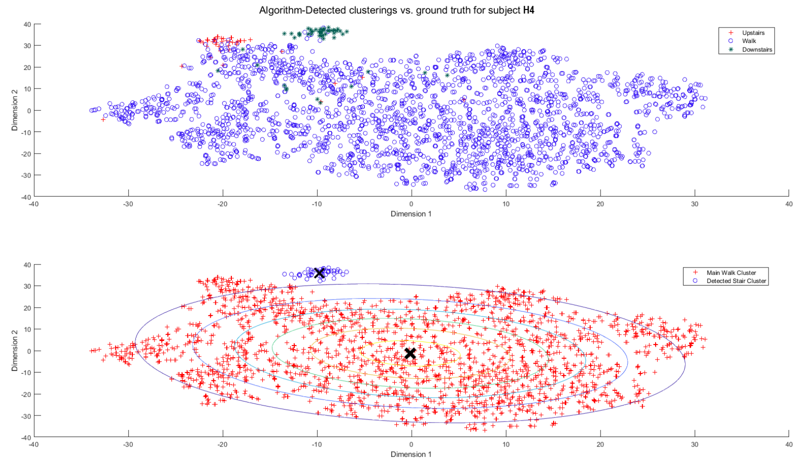


Figure E.3 Top: tSNE cluster model of participant H4, labelled with ground truth activity labels at label resolution 0. || Bottom: tSNE cluster model of participant H4, with activities identified by the proposed clustering algorithm.

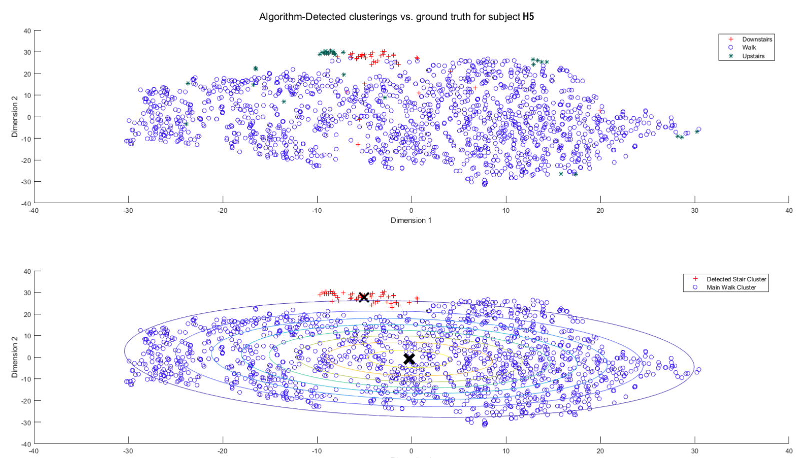


Figure E.4 Top: tSNE cluster model of participant H5, labelled with ground truth activity labels at label resolution 0. || Bottom: tSNE cluster model of participant H5, with activities identified by the proposed clustering algorithm.

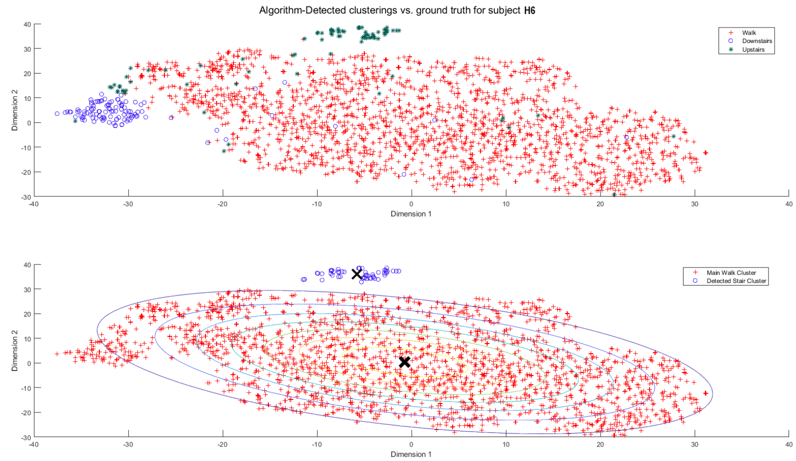


Figure E.5 Top: tSNE cluster model of participant H6, labelled with ground truth activity labels at label resolution 0. || Bottom: tSNE cluster model of participant H6, with activities identified by the proposed clustering algorithm.

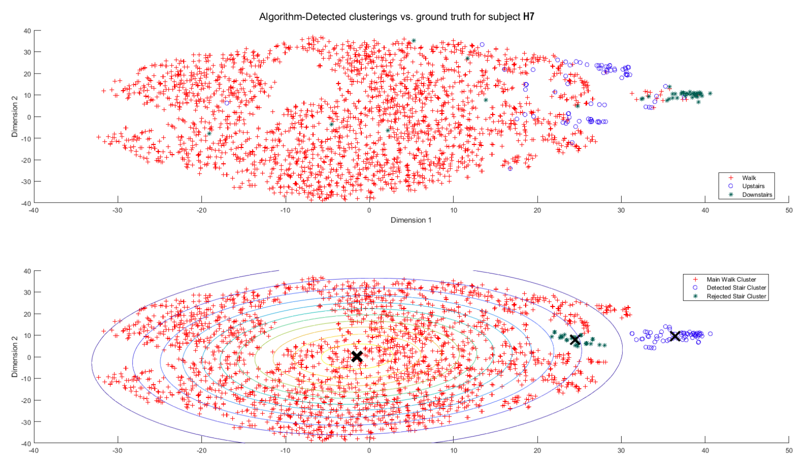


Figure E.6 Top: tSNE cluster model of participant H7, labelled with ground truth activity labels at label resolution 0. || Bottom: tSNE cluster model of participant H7, with activities identified by the proposed clustering algorithm.

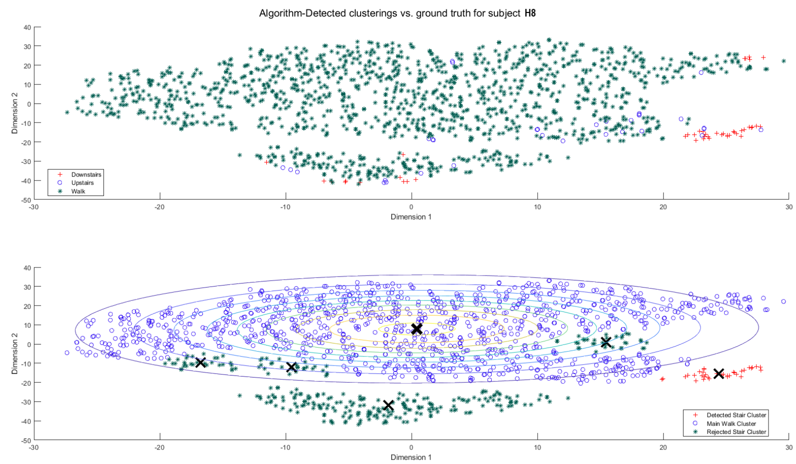


Figure E.7 Top: tSNE cluster model of participant H8, labelled with ground truth activity labels at label resolution 0. || Bottom: tSNE cluster model of participant H8, with activities identified by the proposed clustering algorithm.

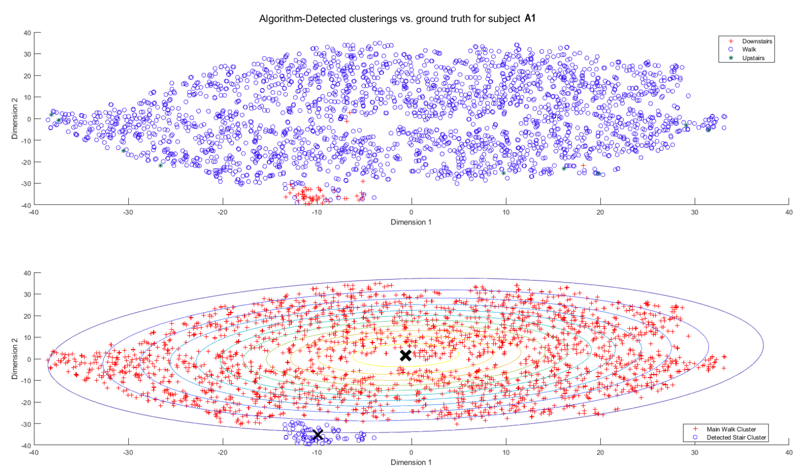


Figure E.8 Top: tSNE cluster model of participant A1, labelled with ground truth activity labels at label resolution 0. || Bottom: tSNE cluster model of participant A1, with activities identified by the proposed clustering algorithm.

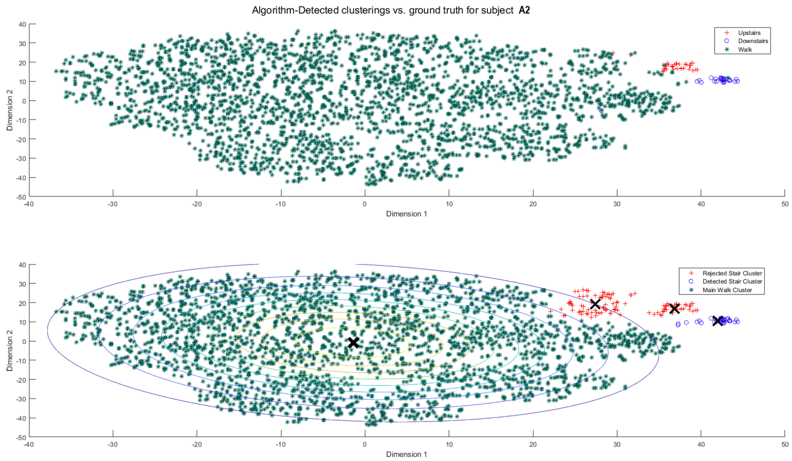


Figure E.9 Top: tSNE cluster model of participant A2, labelled with ground truth activity labels at label resolution 0. || Bottom: tSNE cluster model of participant A2, with activities identified by the proposed clustering algorithm.

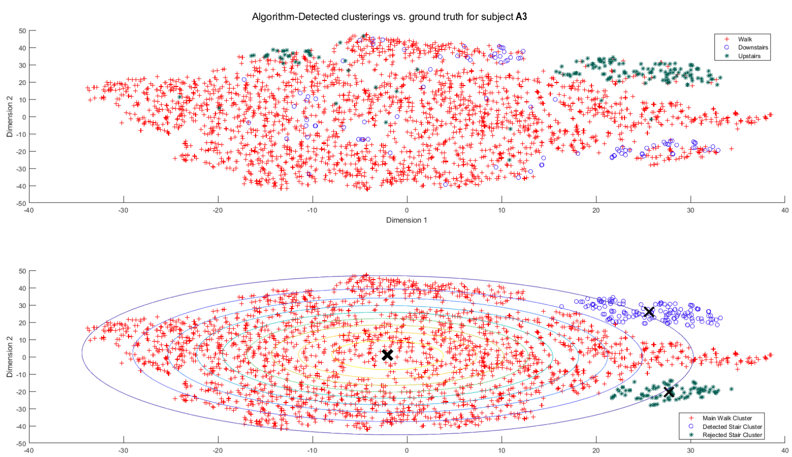


Figure E.10 Top: tSNE cluster model of participant A3, labelled with ground truth activity labels at label resolution 0. || Bottom: tSNE cluster model of participant A3, with activities identified by the proposed clustering algorithm.

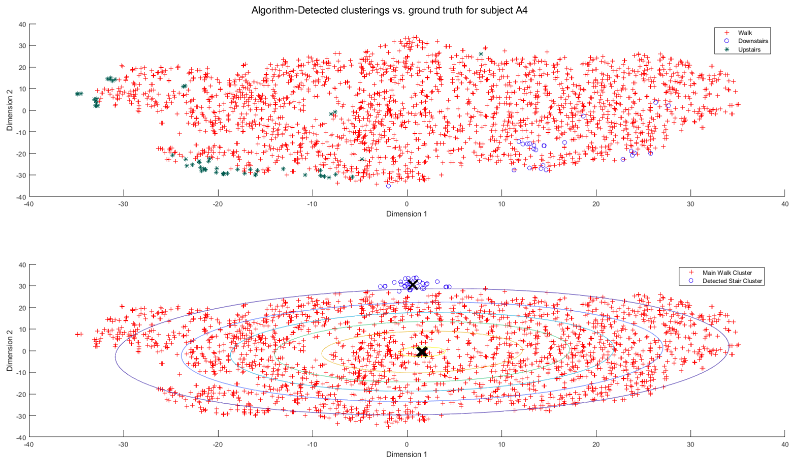


Figure E.11 Top: tSNE cluster model of participant A4, labelled with ground truth activity labels at label resolution 0. || Bottom: tSNE cluster model of participant A4, with activities identified by the proposed clustering algorithm.