ADVANCING PROSTHETIC PRESCRIPTIONS: DEVELOPING A SENSOR-BASED SYSTEM FOR OBJECTIVE ACTIVITY ASSESSMENT

Matthew Wassall

Supervisor: Dr. S. Thies
Prof. M. Granat
Sir S. Zahedi

School of Health and Society
University of Salford

This dissertation is submitted for the degree of
Doctor of Philosophy

April 2025

Table of contents

List of figures

List of tables

Acknowledgements

I would like to express my deepest gratitude to those who have supported me throughout my PhD journey. My sincerest thanks go to my supervisors, Dr Sibylle Thies, Professor Malcolm Granat, and Sir Saeed Zahedi. Sibylle, I couldn't have asked for a better primary supervisor. You have been both a guiding force and a steady source of support, giving me the freedom to explore my ideas while encouraging and sometimes demanding me to take necessary breaks. Your trust in my abilities not only with my work but also on your projects have been invaluable to my growth. Malcolm you have always encouragement to seize every opportunity, from setting up ECPO to organizing the conference, has broadened my academic and professional experience in ways I never anticipated. Saeed your vast expertise and connections have been instrumental, always steering my work in the right direction. All three have also been incredibly supportive and accommodating with my dyslexia by proofreading all my work, which I am truly grateful for.

A huge thanks has to go to all the clinicians and prosthetic users that have not just participated but also guided my research. I diffidently couldn't have completed the PhD without you. The expertise and guidance from the clinicians shaped this research. I cannot thank the prosthetic users that gave up their time to participate in my studies enough. You were all willing to help and on top of just conducting the studies the stories and insight you shared have been invaluable to give me a better understanding of prosthetic life. A special thanks to Chantel and Harry in Portsmouth Enablement Centre that helped recruit.

Being part of the Centre for Doctoral Training (CDT) in Prosthetics and Orthotics has been an incredible experience. The CDT offered countless opportunities— the first-year training, summer schools, and the chance to be a part of an amazing team that organize a conference—that have greatly enriched my journey. Beyond the professional development, the friendships I've formed with fellow CDT students have been a source of constant support. I'll always remember our shared experiences, from academic milestones to tackling the 3 Peaks Challenge together.

On support I would not have survived this PhD without the support and friendship of the other students in PO30. Whether it's having lunch together (especially backed potatoes), restocking the biscuit tin, or taking breaks to solve puzzles, these moments of friendship have made the hard days manageable.

Lastly, my heartfelt thanks go to my family, whose unwavering support at every step of this journey has made all the difference. I could not have done this without you.

Acronyms

2MWT 2 Minute Walk Test

6MWT 6 Minute Walk Test

AALQ Attitude to Artificial Limb Questionnaire

AB AdaBoost

ABC Activities-specific Balance Confidence

ABIS Amputation Body Image Scale

ADAPT Assessment of Daily Activity Performance in Transfemoral Amputees

AGM Acceleration, Gyroscope and magnetometer data

AGMF Acceleration, Gyroscope, magnetometer and free acceleration data

AGMFC Acceleration, Gyroscope, magnetometer, free acceleration and cadence data

AGV accelerations, gyroscope and velocity

AGVC accelerations, gyroscope, velocity and cadence

AMPnoPRO Amputee mobility predictor without prosthesis

AMPro Amputee mobility predictor with prosthesis

AMPROM AMPutee Reported Outcome Measures

AMPSIMM Amputee Single Item Mobility Measure

ANN Artificial Neural Networks

ANOVA Analysis of Variance

ASS Amputee Activity Survey

BBN Bayesian neural networks

BBS Berg Balance Scale

BIQ Body Image Questionnaire

COMPASS Consensus Outcome Measures for Prosthetic and Amputation Services

CNN Convolutional Neural Networks

DBN Dynamic Bayesian Network

DT Decision tree

EMG Electromyographs

FAC Functional Ambulation Categories

FAI Frenchay Activities Index

FAM Functional Assessment Measure

FIM Functional Independence Measure

FMA Functional measure for amputees

FMG Forcemyographs

FNN Feedforward Neural Network

FSST Four Square Step Test

GDPR General Data Protection Regulation

GMM Gaussian mixture model

GPS Global Positioning System

HMM Hidden Markov Model

HSVM Hierarchical Support Vector Machine

IMU Inertial Measurement Unit

ISO International Organization for Standardization

ISPO International Society for Prosthetics and Orthotics

KNN K-Nearest Neighbours

LCI Locomotor Capabilities Index

LDA Linear Discriminant Analysis

LEAD Lower Extremity Amputation Data Set

LEMOCOT Lower-Extremity Motor Coordination Test

LR Logistic Regression

LSTM Long-Short Term Memory neural network

M Mean

MAE Mean Absolute Error

mRMR Minimum Redundancy Maximum Relevance

MS Multiple sclerosis

NB Naïve-Bayes

NCA Neighborhood component analysis

NHS National Health Service

NNR Neural Network Regression

NQ-ACGC Quality of Life in Neurological Conditions – Applied Cognition/General Concerns

OPCS Office of Population Censuses and Surveys Scale

OPUS Orthotics Prosthetics Users Survey

OS Other shank

PEQ Prosthesis Evaluation Questionnaire

PEQ-MS Prosthetic Evaluation Questionnaire, Mobility Subscale

PFI Physical Function Index

PGI Patient Generated Index

PLUS-M Prosthetic Limb Users Survey of Mobility

PPA Prosthetic Profile of the Amputee

PROMIS-29 Patient-Reported Outcomes Measurement Information System 29-Item Profile

PROS Prosthetist's Perception of Client's Ambulatory Abilities

PS Prosthetic Shank

PSFS Patient-Specific Functional Scale

PSSS Perceived Social Stigma Scale

Q-TFA Questionnaire for Persons with a Transfemoral Amputation

QDA Quadratic Discriminant Analysis

RF Random Forest

RFE Recursive Feature Elimination

RMI Rivermead Mobility Index

RP Raw and Person combined

SACH Solid Ankle Cushioned Heel

SAT-PRO Satisfaction with Prosthesis Questionnaire

SCS Socket Comfort Score

SD Standard Deviation

SEW Symmetry of External Work

SF Short Form Health Surveys

SIGAM Special Interest Group of Amputation Medicine

SIP-PD Sickness Impact Profile-Physical Dimension

SMD Standardized Mean Difference

SVM Support Vector Machine

TAPES Trinity Amputation and Prosthesis Experience Scales

TF Transfemoral

TFP Transfemoral Fitting Predictor

TH Thigh

TR Trunk

TT Transtibial

TUG Timed Up and Go

TWT Timed walk test

UK United Kingdom

USA United States of America

WHOQOL-BREF World Health Organization Quality-of-Life Scale – Brief Version

Abstract

The prosthesis components a lower limb amputee receives are determined by their assigned K level. K levels range from K0 to K4 and are defined by the user's ability to traverse environmental barriers, change cadence and ambulation skill. K levels are assessed during a single clinic visit which varies between clinics but mostly utilizes conversation and basic mobility assessments. There are known issues with the reliability of K level assignment, especially when deciding between a K2 and K3. It has been shown that if a lower limb prosthetic user is not given an adequate prosthetic that meets their activity needs it could lead to the patient becoming less active and/or not using their prothesis. This PhD aims to create a sensor-based system to assess a patient's activity levels in the real-world to reduce the issues with reliability during K level assignment.

As a first step, to fully understand the requirements of the system, a study was carried out where interviews were conducted with clinical experts. The ability of the patient to vary their cadence, traverse different terrain, walk without a walking aid and also the distance they can walk were emphasised by clinicians as the main differences between a K2 and K3 patient, and would constitute the data that would be required from the proposed sensor-based system (presently these are only be assessed via self-report).Of these measures only cadence is specifically stated in the current K level definitions, which suggests that the current K levels definitions do not meet the clinical needs.

A review was then conducted to identify the specification of the system in terms of algorithms and sensors that can be used to provide the required data. The review found that cadence has previously been measured with a shank mounted IMU. Moreover, body-worn IMUs have been used to accurately identify between flat ground, stairs, and ramps. However, walking on uneven terrain has not previously been classified for amputee gait. Furthermore, no studies could be found that identified walking aid use using appropriate body or prostheses mounted sensors. The review also looked at what method would be best to process and analyse the data. It was found that K-nearest neighbour, support vector machines, random forest and long short term memory neural networks are classifiers that have previously shown success with similar problems. Using a low pass filter and breaking the data into windows has also shown to be beneficial.

As a second step, a study was conducted to inform further system development, split into two parts. The first part was concerned with collection of data from lower limb prosthesis users in supervised real-world conditions. For this the participants had sensors attached and then were asked to traverse a range of set terrains, with and without a walking aid, outdoors. These data were then been used to train the classifiers. It was found that the terrain a lower limb prosthetic user is traversing can be classified using a single IMU mounted on the prosthetic shank, but walking aid use cannot be classified to clinically acceptable accuracies using a single IMU. A random forest model produced the highest terrain classification accuracies.

The second part of the study was conducted in a gait lab. For this part the participants were asked to traverse similar terrains, with and without a walking aid, but with sensor and full motion capture data being collected. These data have been used to create virtual sensors that were then used to estimate the ideal IMU location to increase classification accuracy. Feature importance was used to identify the most important aspects of the accelerations and then variations in these parts of the acceleration data were examined for different locations on the prosthetic shank. It was discovered that a consistent location is critical for high classification accuracies, and that for terrain classification accelerations captured at the ankle produce higher accuracies.

A final study was conducted to explore clinical experts' views on the developed system and the output data. Real-world data was collected from 3 lower limb prosthetic users over two weeks using a prosthesis shank-worn IMU. These data were processed and classified using the previously created algorithms, to estimate terrain and walking aid use. Each participant also had their K level clinically assessed using current standard clinical procedures. A report was compiled for each participant that summarised the clinical assessments and the classification data. These reports were shown to 4 clinical experts and semi-structured interviews were conducted to assess their thoughts on the data, if the system would be clinically useful and if the data would change their K level assessments. All the interviewees thought the data would help with clinical assessments and had positive views about the data that were produced. They also all commented on how the data would change how they would conduct a K level assessment and two said that for two of the participants the sensor data would change the K level they would assign compared to just the clinical assessment. Active time and load through the

prosthetic were the only measures that were identified that could further improve the system for clinical use.

Chapter 1

Background and introduction

1.1 Background

1.1.1 Lower limb prosthetic use

It is estimated that 35–40 million people globally require prosthetics and orthotics [9] and the need for prosthesis is on the rise, for example in 2005 there were 1.6 million people living with limb loss in the United States of America (USA) and it is projected that there will be 3.6 million by 2050 [10]; this is a realistic estimation as in 2021 there were 2.2 million people with limb loss in the USA [11]. Underlying reasons for these numbers include that over half a million individuals experience limb loss or are born with limb difference in the USA each year, with 83% of limb loss being lower body [11]. 86.4% of those were above 45 years old, with 44.7% above 65, 57.6% were due to diabetes and only 12.9% due to trauma. In the United Kingdom (UK) limb loss statistics are similar with 70% of lower limb prosthetic users being over 54 years old, with only 10% for trauma [12].

Diabetes mellitus, commonly referred to as diabetes, is a chronic condition that occurs when raised levels of blood glucose occur because the body cannot produce any or enough of the hormone insulin or cannot effectively use the insulin it produces [13]. Diabetes can often lead to lower limb loss as amputation may become a necessity when blood flow to the extremities is reduced due to diabetic peripheral arterial disease [14]. It has been estimated that 537 million people worldwide have diabetes and that this will increase to 783 million by 2045 [13]. It was estimated that over 6.7 million people aged 20-79 will die from diabetes related causes in 2021, and that the direct healthcare costs due to diabetes will exceed one trillion USD by 2030 [13]. Over 90% of cases of diabetes worldwide are of Type 2 diabetes. The causes of Type 2 diabetes are not known but there are strong links to weight, age and ethnicity [13]. One of the key aspects to management of Type 2 diabetes is promoting a healthy lifestyle that includes regular physical activity [15].

Lower limb prosthetic users have been found to be less active than non-amputees [16], with research suggesting that a large proportion could be classified as sedentary [17]. Inactivity has been shown to be a serious cause of death with an estimated 1.9 million deaths globally in 2002 linked to inactivity [18]. In the USA it was found that people who have limb loss have a mortality rate 18.5% higher than the general population [11].

When limb loss occurs, prosthetics are generally prescribed to enhance mobility and physical functioning. Depending on the level of limb loss, a transtibial (TT) or transfemoral (TF) prosthetic may be prescribed for the lower limb. Transtibial prosthetic users make up 54% of the prosthetic users in the UK, with Transfemoral accounting for 37% [12]. The main components of a lower limb prosthetic are: socket, suspension mechanism, knee joint (above knee limb loss only), pylon, and ankle and foot joint [19].

Given the large number of individuals experiencing lower limb loss, prosthetic devices play a crucial role in restoring mobility and improving quality of life. Various types of prosthetic knee joints and components are available to accommodate different levels of activity and user needs. There are two main categories of prosthetic knee joints, mechanical and microprocessor. Mechanical knees can be single axis, locking, polycentric or hydraulic. Single axis knees are considered the most basic and bend freely on a single axis, that acts like a hinge, and do not have any stance control, which can be an issue for older users [20]. They are lightweight and low cost [21]. Locking knees are similar to single axis knees but have a locking feature to keep the knee straight when in stance, this can be manually controlled or weight activated [20]. Polycentric knees have multiple axis of rotation, usually utilising a 4-bar link mechanism. These offer more stability in stance and due to the mechanism the limb length shortens during swing which increases foot clearance [22]. Hydraulic knees are incorporated into a single or polycentric knee to control the swing phase of a stride through hydraulic dampening [20]. This gives the user the ability to vary walking speeds as the knee will limit fluid flow and therefore limit flexion [21]. These components add cost and weight to the prosthesis [22]. Microprocessor knees utilise sensors in the knee that then influences the response on the knee similar to a hydraulic knee [21]. With the quick response and adaptability of the system during a stride, they can produce a more natural gait [22]. These are much more expensive and heavier than other knees [22]. For an example of the weight difference between different knees, an

Össur Locking Knee weighs 0.248Kg [23] and an Össur Power Knee weighs 2.6Kg [24].

There are also different types of ankle and foot joints are: solid ankle cushioned heel (SACH), flexible keel, axial, dynamic response, and microprocessor [25]. A SACH foot is the most basic prosthetic ankle and foot, it is a non-articulated foot with no hinged parts [26]. They are low cost and durable but the heel height is fixed and they have poor toe-off [27]. Flexible keel feet are similar to a SACH foot but with an elastic keel [25], meaning the forefoot is able to conform to uneven terrain [26]. Axial feet can be single or multi-axis. Single axis allows for dorsiflexion and plantarflexion, where as a multi-axis foot also allows medial lateral flexion [25]. The ankle movement helps to dampen some of the stresses of walking which can reduce pressure in the prosthetic socket [27]. A dynamic response foot stores and releases energy during a stride, absorbing energy in the keel during the roll-over phase and then springs back to provide a subjective sense of push-off [26], but due to the ankles inability to produce a plantar-flexion moment when in a plantar-flexed position they cannot produce a natural push-off at toe-off [28]. This allows for a more natural gait [25]. Microprocessor feet are similar to microprocessor knees in that they utilise sensors in the foot to make automatic adjustments so the foot can adapt to the terrain. They also incorporate ankle motors to give an active push at toe-off to help provide a more natural gait, but they are high cost and heavier than other prosthetic feet [27]. For an example of the different weights of different types of feet, an Ottobock Kintrol single axis foot weighs 0.953Kg and an Ottobock Empower microprocessor foot weighs 2.145Kg, both for 27cm feet.

Baars et al. [29] conducted a review of satisfaction with lower limb prosthetics, they found that users only had a mean satisfaction score of 58 (range 0-100) in relation to prosthetic weight and that TF prosthetic users were more likely to be dissatisfied with their prosthesis than a TT prosthetic user.

In summary, lower limb loss is a growing global health issue, with diabetes as a primary contributor. Lower limb prostheses for TT and TF patients offer a solution to aid with mobility, and various types exist that differ in components, general functioning and also weight and costs. However, satisfaction with lower limb prostheses remains mixed.

1.1.2 K levels and associated prosthesis component prescription

K levels

The prosthetic components a lower limb prosthetic user is given can have an impact on the types of activities the user can do. As such, if a lower limb prosthetic user is not given an adequate prosthetic that meets their activity needs it could lead to the patient becoming less active and/or not using their prothesis [5], and the implications of that on health have been discussed above. Activity levels of lower limb prosthesis patients are characterized by K levels. A K level is assigned to the patient, and the K level designates the prosthesis components the patient is given and the payments from their medical insurance that they will receive [30]. K levels were created by the US Health Care Financing Administration in their common procedure coding system [31]. K levels go from K0 to K4, and a description of each level is provided in Table 1.1 below. These brief descriptions are the only definitions given to distinguish between the different K levels with not additional information given. There is no guidance or objective measures to define any aspect of the definitions, for example how much variance in cadence would constitute being able to ambulation with variable cadence. This brings subjectivity into assigning K levels which could mean different clinicians judge the same level of an aspect of activity to align to a different K level. There is also no set procedure to assess patients for classification of K levels. Balk et al. [5] reviewed studies that looked at the validity of 50 outcome measures that could be used to classify a patient's activity level for component prescriptions A.1. All the outcome measures were deemed to have some validity, but there remains little evidence to show treatment effects, patient satisfaction and long-term use effects. Balk et al. stated that current research fails to address whether different lower limb prosthetic users will benefit from specific prosthetic components. It was also found that there is sparse evidence of long-term outcomes. They also stated that no study had investigated prosthetic prescriptions in relation to functional outcomes, and that K level assessments variability and subjectivity could lead to inappropriate prosthetic prescriptions. Following this, it was suggested that more research is needed to address the issue of heterogeneity of treatment effects and improve matching of components to patients. Hafner [32] stated the same findings already in 2005 which highlights the remaining lack of research in this area to address these issues.

K-levels are crucial in determining the appropriate prosthetic components prescribed, as they classify patients based on activity levels and predict potential mobility. However,

Table 1.1 K level descriptions [5]

K level	Description
Level 0	Does not have the ability or potential to ambulate or transfer safely with or without assistance and a prosthesis does not enhance their quality of life or mobility.
Level 1	Has the ability or potential to use a prosthesis for transfers or ambulation on level surfaces at fixed cadence. Typical of the limited and unlimited household ambulator.
Level 2	Has the ability or potential for ambulation with the ability to traverse low level environmental barriers such as curbs, stairs, or uneven surfaces. Typical of the limited community ambulator.
Level 3	Has the ability or potential for ambulation with variable cadence. Typical of the community ambulator who has the ability to traverse most environmental barriers and may have vocational, therapeutic, or exercise activity that demands prosthetic utilization beyond simple locomotion.
Level 4	Has the ability or potential for prosthetic ambulation that exceeds basic ambulation skills, exhibiting high impact, stress, or energy levels.

the current K-level assessment process is limited by subjectivity and variability, leading to misclassifications.

Prosthetic component prescriptions

Clinics and manufacturers designate different types of components to different K levels. For example, the NHS requires patients to be a K3 to be able to be given a microprocessor knee. Clinics that publish the data [33][34][35] are in consensus that a K 1 patient would get a SACH foot and single axis manual locking knee if needed (Figure 1.1a, 1.2a). A K2 patient would get a flexible keel foot with a multi axis ankle and a single axis knee with extension assist (Figure 1.1b, 1.2b). A K3 or above patient would get a flex foot, energy storing or shock absorbing foot with a dynamic response or powered ankle and a hydraulic, pneumatic or microprocessor knee (Figure 1.1c, 1.2c).

Fig. 1.1 Ottobock prosthetic feet a) Kintrol K1 foot (www.ottobock.com) [1], b) Terion K2 foot (www.ottobock.com) [2], c) Taleo Harmony K3 foot (www.ottobock.com) [3].

Fig. 1.2 Ossur prosthetic knee joints a) Locking Knee K1, b) Total Knee 1900 K2, c) Power Knee K3 (www.ossur.com) [4]).

Table B.1 in Appendix B.1 displays the various components that Limbs 4 Life recommends for different K levels. Limbs 4 Life is an Australian charity that supports and advises amputees [8]. Training material by Nelson et al. [36] state the same, that a K2 should get a SACH foot and a K4 get a dynamic response foot.

Some evidence exists that K level and associated component selection indeed affect gait. For example, Symmetry of External Work (SEW) is an outcome measure that has shown kinetic gait differences between different types of prosthetic feet. SEW compares the ground reaction force between the prosthetic and intact limb. It has been shown that a K2 and K3 patient have a higher SEW using a K3 foot over a K2 foot for ramp ascending and descending. Interestingly, there was no statistical difference for a K2 patient in using a microprocessor-controlled ankle and a K2 foot, whereas for a K3 patient there was a large improvement for the microprocessor-controlled ankle and a K2 foot [37]. This suggests that more active patients will benefit more from using more advance components and are more capable of using more advanced components, hence appropriate K level assignment is important to optimize the match between patient and prosthesis components.

1.1.3 Problems with K level assignment

Evidence exists that mis-classification of K levels at times occurs. For example, data collected at Blatchford centres, for ISO standards, published at ISPO, present the case of a 72-year-old man that was classed as K2 but after 8 months was walking over 6000 steps a week and wearing his prosthesis 12 hours a day, so arguably he should have been classified as a K3 level prosthetic user [38]. The paper also stated that, "The K level will NOT Select Best Prescription for amputee". Furthermore, Amputee Reported Outcome Measures (AMPROM) data obtained from the online AMPROM database highlight difficulties with the assignment of K levels, see Table 1.2. AMPROM is a database set up by ISPO UK to collect data on lower limb prosthetic user outcome measures [6]. Only a small number of prosthesis users data had been recorded in this database at the time of writing this thesis, but this is the only database available that records this type of data for prosthetic users. The data show K2 and K3 patients classified by the same clinician between July 2018 and July 2020. All of the recoded prosthetic users except two used a walking aid, regardless of K level, and type of walking aid varied across participants. Some K2 patients covered more distance in the 6-minute walk test, used less walking aids and stumbled and fell less often than some K3 patients. Whilst some outcomes rely on the accurate report of the patient, this table overall highlights the difficulty of distinguishing between K levels, in particular between K2 and K3.

Table 1.2 Data from AMPROM, ISPO UK NMS., retrieved on 8th of January 2021 at 16:00pm, URL: https://amprom.uk [6]

Level/s of amputation	K-level	Walking aids	Walking aids - Usage	Distance covered in 6 minutes (m)	How frequently do you stumble?	How frequently do you fall?
RIGHT Transfemoral	K3	no information	Outdoors only	430	Less than once monthly	Less than once monthly
RIGHT Transfemoral	K3	no information	Outdoors only	430	Once or more monthly	Once or more monthly
RIGHT Transtibial	K2	RIGHT Cane	Outdoors only	400	Once or more monthly	Once or more weekly
RIGHT Transtibial	K2	RIGHT Cane	Outdoors only	400	Once or more monthly	Once or more weekly
RIGHT Transfemoral	K3	RIGHT Cane	Outdoors only	208	More than once daily	Once or more weekly
RIGHT Transfemoral	K2	Crutches	All the time	200	Less than once monthly	Less than once monthly
RIGHT Transfemoral	K2	Crutches	All the time	200	Less than once monthly	Less than once monthly
Right Symes	K2	Walker	Outdoors only	150	Once or more monthly	Once daily
LEFT Transfemoral	K3	LEFT Cane	Outdoors only	128	Less than once monthly	Less than once monthly
RIGHT Knee	K3	Walker	All the time	100	More than once daily	More than once daily
RIGHT Toe	K2	RIGHT Cane	no information	67	Less than once monthly	More than once daily

1.1.4 Current clinical practice for assessment of K levels

In line with the above, a systematic review [39] identified the need to establish detailed outcome measures to assess lower limb prosthetic users' activity levels. Subsequently, Jamieson [40] ran focus groups and interviews with lower limb prosthetic users and clinicians to investigate their views on prosthetic rehabilitation outcome measures and explore how an activity monitoring system could be beneficial. The research found that activity monitoring was not currently used in assessing patient activity levels, patients do not keep their clinicians updated on their activity levels except when needed and measures on the terrain that the patients are traversing would be beneficial for clinical decision making.

Arguably, a vast range of outcome measures have been developed to assess different aspects of a patient's health and activities. Table 1.3 displays the main ones that are clinically used; a review was conducted [7] that assessed most of the outcome measures listed. The review concluded that even for outcome measures that are valid and reliable, there remain issues with ease of use. It was also highlighted that there is no gold standard measure and no guidance as to which measures should be used when allocating K levels and prescribing components to a patient. Importantly, most measures identified were self-reported, yet studies have found that self-reported measures are not accurate when compared to objective measures [41][42], for example, it has been shown with some non-prosthetic users, that timed walking tests do not correlate with daily stepping activities [43].

A more recent and comprehensive review by ISPO as part of the LEAD and COMPASS project found 60 outcome measures for lower limb prosthetic users [44]. This project then had an expert panel examine the outcome measures and a list of recommended outcome measure sets were presented. Three sets were recommended: COMPASS – Time-up-and-go test, Amputee Mobility Predictor, 2-Minute Walk Test, Prosthesis Evaluation Questionnaire Residual Limb Health and Utility sections, and Trinity Amputation and Prosthesis Experience Scales-Revised: COMPASS+ - 6-minute walk test and Comprehensive High-Level Activity Mobility Predictor; COMPASS Adjunct - Patient Specific Function Scale. The COMPASS set is recommended as the standard set and takes approximately 45 minutes to complete. The COMPASS+ is for high-functioning individuals and takes approximately 25 minutes to complete. COMPASS Adjunct is recommended to be used in addition to the other sets for a clinical setting and takes approximately 10 minutes to conduct. No published papers that could be found that investigated the use of any of these recommended outcome measure sets for K level assignment or prosthetic prescriptions; the ISPO report also does not mention whether these outcome measures could be used for lower limb prosthetic prescriptions.

When this project started, notably no work could be found that looked at gaining an understanding of how clinicians view K levels and how K level assessments are conducted in different centres. A vast number of clinical outcome measures exist (see Table 1.4), yet understanding their implementation (or lack of such) and associated clinical views and processes may help identify the sources of the inaccuracies and have the potential to inform design of an assessment method that addresses present shortfalls. Activity monitoring through small body-worn devices may be a solution, as it has shown promise in the research arena as will be discussed in the following sections.

In summary K levels are fundamental to prosthetic care, determining the prescription a user will be given and the benefits they will receive, but also prosthetic components are designed to specific K levels. Even so, with this there is evidence of inaccuracies in K level assignments, which could be due to the subjectivity of the assessments and the vague definitions. There are validated outcome measures that can be used to assess lower limb prosthetic users but there is no recommendation on how these should influence K level assignment. The prosthetic component a lower limb prosthetic user is give can have a large impact on their activity levels and consequently their health and wellbeing, therefore it is vital that these issues with K level assignment are investigated and reduced. So a gap remains in developing objective, standardized assessment methods that could improve the accuracy and consistency of K level assignments and better match prosthetic components to patient needs. This demonstrates the need for exploring objective, real-world data collection and monitoring methods to enhance K level accuracy and improve patient outcomes.

Table 1.3 Outcome measures analysis. List of measures from Condie et al. [7]

Outcome measure	Description	Advantages	Disadvantages
Socket comfort score (SCS)	Patients are asked to rate the comfort of their socket from 0 to 10 scale where 0 is the most uncomfortable and 10 is the most comfortable [45]	It has been shown that reported SCS was consistent and reliable, with a strong relationship between SCS and clinical evidence of poor fit [45]	Only assesses the comfort of the socket and not the ability of the patients. [7]
Trinity Amputation and Prosthesis Experience Scales (TAPES)	TAPES is a self-administered questionnaire. It consists of nine subscales. 3 regarding psychosocial, 3 covering activity restriction and 3 additional that assess satisfaction with the prosthesis [46].	TAPES has shown good internal consistency [7]	TAPES is intended to assess the adjustment to amputation and not assess the ability of the patient [7]
Perceived Social Stigma Scale (PSSS)	PSS is a self-reported questionnaire the explores quality of life in respect to social stigmas about limb loss. It consists of 22 items derived from a large pool of attributes that embody common negative stereotypes associated with disabled people [7]	The internal consistency of the PSSS is good, and it is reported to have good face and content validity [7].	PSSS dose not assess a patient's ability but how they are affected by limb loss [7]
Prosthesis Evaluation Questionnaire (PEQ)	PEQ is a self-reported questionnaire that aims to measure functional outcomes of lower limb prostheses. The questions refer to the 4 weeks immediately preceding the administration of the questionnaire. Questions relate to ambulation, limb health, appearance and quality of life. [47]	Testing has shown the PEQ to have good reliability and good to excellent construct validity [7]	

Body Image Questionnaire (BIQ)	BIQ is a self-reported questionnaire that measures quality of life. It asks respondents about their feelings about their body shape, the shape of their prostheses, the attitude of others towards them, and the impact on their social activity [7]).	The BIQ has good internal consistency [7]	Aspects of reliability and validity are unknown and the BIQ does not assess the patient's ability [7]
Amputation Body Image Scale (ABIS)	ABIS is a self-reported survey that assess how a person with an amputation perceives and feels about their body. [48]	Correlation to the TAPES has been shown and good reliability demonstrated [7]	ABIS does not assess a patient's ability [7]
Attitude to Artificial Limb Questionnaire (AALQ)	AALQ is a self-reported questionnaire that measures quality of life. It measures satisfaction with prosthesis, walking ability, attitude of others to them, and restoration of body image [49].	Internal consistency of the AALQ is good [7]	Some psychometric properties remain unreported [7]
Frenchay activities index (FAI)	FAI is a self-reported questionnaire that measures activities of daily living. It has 3 domains which are domestic chores, leisure/work and outdoor activities [7]	The FAI is reported to have excellent reliability and adequate validity [50]	FAI looks at the frequency of an activity not the ability to do an activity [7]
Prosthetic profile of amputee (PPA)	PPA is a self-reported questionnaire that measures the factors related to prosthetic use and the use of the prosthesis. [51]	The PPA has been shown to be valid and reliable for clinical use [7]	Problems have been reported with self-administration and patient's understanding of the questions [7]

Houghton scale	Houghton scale is a self-reported questionnaire that measures the wear and use of a prosthesis. It has been proposed that the Houghton scale parallels K levels [52]. It consists of four items: the amount of time the prosthesis is used, the manner in which it is used, whether an assistive device is used outside, and the individual's perception of stability while walking outside on a variety of terrain [7].	It has been shown to be responsive to change in prosthetic use [52]	Some floor and ceiling effects have been reported [7]
Amputee Activity Score (AAS)	ASS is a self-reported questionnaire which explores activities of daily living and the frequency of participation in certain activities [53].	AAS has been shown to be responsive to change in mobility with rehabilitation and at follow-up [7]	Construct validity is still unclear [7]
Special interest group in amputee medicine (SIGAM)	SIGAM is a self-reported questionnaire that assesses a patient's mobility. It asks questions about walking ability and grades patients from A to F [54].	It has been shown to be reliable and responsive to change in mobility [7]	More focused on low mobility patients [7]
Locomotor capabilities index (LCI)	LCI is a self-reported questionnaire that asks the patients to assess if and how they complete certain activities [55].	LCI has been shown to have good internal consistency, test-retest reliability and construct validity [7]	The activities assessed are limited and may not fully assess a patient's ability [7]
Functional measure for amputees (FMA)	FMA is a self-reported questionnaire that is a reduced version of the PPA, which was deemed to take too long to complete [56].	FMA has been shown to have good reliability [7]	FMA has been shown to have poor validity and the activities assessed are limited and may not fully assess a patient's ability [7]

Amputee mobility predictor with (AMPPro) / without prosthesis, (AMPnoPRO)	AMPPro and AMPnoPro are a set of objective tests that are used to measure a patients mobility level. It covers sitting balance, transfers, standing balance, gait, stairs, and use of an assistive device. Score ranges have been allocated to different K levels [57].	The AMP with and without a prosthesis are reliable and valid measures for the assessment of functional ambulation in lower-limb amputee subjects [57]	
Timed walk test (TWT)	Timing of walking can be carried out by either testing speed over a short distance (e.g., 10 metres) or endurance in which the subject is asked to walk as far as they can in a given time (i.e 6 minutes) [7].	TWTs are known to be valid and reliable with a variety of clinical conditions and are frequently used as the gold standard comparator test [7]	Poor construct validity has been shown depending on comparison measure [7]
Timed get up and go	The subject is observed rising from an armchair, walking 3 m, and returning to the chair on a standard carpet (Sears, 2012).	The Timed up and go test showed good intra-rater and inter-rater reliability [7]	Timed get up and go test only assesses a limited element of a patient's mobility [7]

1.2 Review of past studies using activity monitoring

This section looks at activity monitoring for healthcare. Activity monitoring can be used to collect data on a person's activities in the real-world, which could have benefits for lower limb prosthetic prescriptions as it would give clinicians more information on their patients to make a more informed decision without taking up additional clinical time. Activity monitors can record cadence and therefore cadence variability which is stated in the K level definitions as a distinguishing measure between K2 and K3. This section will look at research relating to activity monitoring of prosthetic users first to establish relevant context, and then non-prosthetic users to explore broader monitoring approaches that may offer transferable insights.

1.2.1 Activity classification in prosthesis users

Chadwell et al. [58] performed a systematic review of activity monitors used in studies related to prothesis users, as activity monitoring was deemed as a potential method of understanding how prostheses are used in everyday life. The studies reviewed looked at validating systems or algorithms, comparing activity measures to clinical assessments, seeing how interventions changed activity levels, and comparing different populations' activity levels. The studies that attempted to validate new systems and the ones that compared activity measures to clinical scores are the most relevant for this research. The majority of the validation studies were looking to validate a body-worn, sensor-based system to monitor activities, but there were some that explored different approaches, like using GPS or smartphone sensors. The most common clinical score compared to everyday activity levels of lower limb prothesis patients was K levels. A few studies found a good correlation between K level and number of steps taken [59][60][61][62][63][64]. However, only 3 studies measured data for longer than three months with most recording for less than 2 weeks. Nevertheless, the review concluded that monitoring patients outside of the clinic can give a clearer picture of the patient's capabilities and requirements. A more detailed discussion of the most relevant studies reviewed is provided below.

Use of accelerometers to monitor aspects of activity.

Several studies concerned with validation used accelerometers attached to a prosthetic user's body or prosthesis to measure step count, walking velocity, step length and to classify locomotion [65][66][67]. Notably, Redfield et al. [67] found that with one accelerometer attached to the ankle of the prosthesis 90.1% accuracy could be achieved when identifying sitting, standing, moving, and doffing (putting on a prosthesis). Gardner et al. [65] also showed that using a proximity sensor in the cuff of the prosthesis could detect doffing. Gardner et al. [65] used two accelerometers with one attached to the thigh and one to the prosthetic ankle. However, it may be that some patients may not be willing to attach a sensor to their body for longer time periods.

One study compared measured to subjectively-reported activity: Balkman et al. [42] also used accelerometers and looked at how prosthetists' and patients' view of the patient's activity level compared to objective data. The study found that none of the prosthetists' estimates for wear time matched the measured data, and only one of the patients' self-evaluations was predicted accurately. The estimated values for sitting, standing, and walking time showed no correlation to the values measured, hence the authors rightly questioned the reliability and accuracy of subjective estimates of activity levels as are presently used for K level classification.

Three further studies compared K levels to measurements of activity. One study by Arch et al. [59] looked at cadence of the participants to distinguish between K2 and K3 patients. Their study found that on average K2 patients walked at a slower cadence and with less variability in their cadence. However, almost 15% of participants were found to show activity characteristics that were of a different K level to the one they had initially been classified as. The measurements were made using a wrist-worn FitBit; due to using a wrist-worn device there might be some error in the cadence measurements but such placement would make it easier to get the participants to wear the device, for example hidden in a watch or armband. Arch et al. also used a FitBit to compare activity data to Medicare functional classification levels, which are the same as K levels [68]. The clinical assessment consisted of a ten-metre walk test and a six-minute walk test. On average K2 patients were found to have a slower walking speed, lower step count, shorter distance covered in the six-minute walk test and fewer active minutes than the average K3 patient. Just over 18% of participants aligned with the average of a different K level, and not the average of their assigned K level, in terms of active minutes, total steps, walking speed and percentage of time conducting high level activities. A waist-worn FitBit was used by Albert et al. [61] to classify activity levels and compare them to K levels. The study recruited nine transfemoral prosthetic users but seven were assessed as K3, with one assessed K2 and one K4. The activity levels were classified as light, fairly and very active; the K2 participant did have the highest percentage of their activity as light but this was only slightly higher than two of the K3 participants, and the K2 user did not have any very active periods, but this was also the case for three K3 participants. The K4 participant did have the highest percentage of their activity as very active but had a larger percentage of light activity than one K3 participant. Due to there only being one K2 and K4 participant each, this can not be accepted as a validated outcome. The other study by Orendurff et al. [64] investigated the relationship between K levels and real-world ambulatory activity data. An estimated K level was calculated from the measured data and compared to the patient's clinically-assessed K level. Calculated levels correlated with clinically assessed K levels, however, the K2 patients scored similar to the K3 patients. The estimated K levels rated potential to ambulate, cadence variability and energy level for each participant on a scale between 0 and 4.9. The average of these ratings combined with the participants clinically assigned K level was used as the estimated K level. This means that the participants clinically assigned K level had a 25% influence over the estimated K level and therefore this study does not directly compare activity levels to K levels. If the assigned K level was taken out of the calculation, there would be one of the three K2 patients that would have been more active than two of the six K3 patients, clearly demonstrating challenges with assigning K2/K3 levels. The same calculated K level was also used by Orendurff et al. in a study that compared this measure against clinician-assessed K levels, but the clinicians were only shown step count and cadence data from the prosthetic users [63]. There was a strong correlation between the k-levels derived using the 2 different approaches (R^2=0.829), but the clinician-assessed K levels were purely assigned from step count and cadence data and therefore cannot be accepted as the same as clinically assigned K levels, because the prosthetists in the study all highlighted other factors that they would use to assign K levels in routine clinical care.

Studies incorporating different sensors.

Regarding some different approaches, one study used GPS data along with a step counter to quantify community mobility and social interactions [69]. It was found that steps in and outside of the home, wheelchair use, prothesis use, driving trips and time spent on social trips could all be quantified. This gives a good picture of the patient's life outside of the clinic and therefore their requirements in terms of prosthesis function. However, the GPS device required the participant to charge it every night and carry it around in a pocket or bag, and this may not be practical for every patient. Another study [62] that incorporated GPS, investigated the Modus Trex derived K level, an estimated measure for K level from StepWatch data although the exact algorithm has not been published, and a modified clinical K level (MCK), derived from criteria set for each K level based on specific measure collected by the GPS and StepWach combined (mean daily step count, peak daily step count, steps in a community setting, peak cadence and environmental barriers traversed), against clinically assess K levels. Out of the twenty-seven participants the Modus Trex K levels were different to the clinically assigned K levels four times (14.8%) and the MCK was different to the clinical K levels 3 times (11.1%).

Two studies used the accelerometers and gyroscopes built into a smartphone. Albert et al. [60] used such to quantify activity levels and compare them with K levels. It was found that there was a correlation between K levels and the proportion of moderate to high level activities. The activities were categorised as either high, medium, low or inactivity, but the type of activity was not identified. Another study used smartphone sensors to detect falls [70]. It was shown that the machine learning algorithms could detect falls for lower limb prothesis users from the mobile phone sensors. Although this is not related to K levels it shows that simple sensors can classify some aspects of human movement.

One other study by Frossard et al. [71] looked at the viability of using a six-axis force transducer integrated into an osseointergrated trans-femoral prosthetic to measure daily activity. The study found that the transducer accurately measured maximum load, gait cycles, cadence, and activity periods. However, the force transducer used required to be integrated into a unique prosthesis shaft which limits its use.

Other studies using activity monitoring on a lower limb prosthetic population. Jamieson [40] looked at developing a system to recognise activities of lower limb prosthetic users in free living conditions. In doing this, a study was run using one thigh mounted accelerometer and activities were classified using a range of machine learning algorithms (Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Random Forest (RF), AdaBoost (AB), Naïve-Bayes (NB), Long-Short Term Memory neural network (LSTM) and Linear Discriminant Analysis (LDA)). The study had eight healthy participants and four transtibial amputees walking on a range of flat ground, stairs, slopes, and soft ground. SVM and LSTM produce the best accuracies of 76.28% and 78.43%, respectively, using a 5-fold cross-validation test but only 56.68% and 31.10% using a leave one subject out approach. This work also looked at unsupervised cluster analysis of the activity data. T-distributed stochastic embedding was deemed to be the best approach and it was found that there were appreciable clusters between ground walking and stair ambulation only. The results of this work show that terrain recognition is possible with lower limb prosthetic users but also that there are limitations in only using one accelerometer mounted on the thigh. The accuracies would have also been affected due to the limited variety in the data collected, with only 131 downstairs samples and 3506 flat ground samples collected.

Accelerometers have also been used to measure gait symmetry [72]. The accelerometer was attached to the participant at the thorax, and the results were found to be acceptable to assess the gait symmetry and regularity of a trans-femoral amputee. The issue with this study's proposed system is the placement of the accelerometer, but it did give an indication on how ambitious the system could be with the amount of data collected from one accelerometer.

In summary activity monitors have shown potential to be powerful tools that could aid in prosthetic prescriptions. A range of sensors recording a variety of measures have been investigated. Accelerometers have been used in many studies, especially for quantifying step count and cadence. There is no consensus, however, on the most effective sensors and measures that would aid in K level assignment.

1.2.2 Activity classification in non-prosthetic users

Research into activity monitoring in healthcare extends well beyond prosthetic users and has produced a wide range of methods. Studies on non-prosthetic users have explored diverse aspects of activity monitoring, from movement classification to rehabilitation tracking, providing valuable methodologies that could be adapted for prosthetic applications. This section will examine a few that have looked at different aspects of activity to present the range of applications that this technology can be utilised for. Camargo et al. [73] showed that you can use Inertial Measurement Units (IMU) and goniometers to calculate walking speed, stair height and ramp incline to a high degree of accuracy. The system used four IMUs, on the foot, shank, thigh, and trunk, and 3 goniometers, on the ankle, knee, and hip. The shank IMU reduced the error the most when calculating the walking speed, while the goniometers reduced the most errors in calculating the step height and ramp angle. Without the goniometers, the error of the step height was about 5 cm, and the error of the ramp incline was about 3.5°, but with the goniometer the errors improved to 1.29 cm and 1.25°, respectively. The shank IMU could detect a ramp but could not accurately measure the incline. The classifier proving most successful in terms of reducing errors was a feedforward neural network with a Kalman filter. However, goniometers measure angles, hence, will not be effective on a transtibial prosthesis.

Other studies have used machine learning algorithms to classify movements. Chen et al. [74] found that K nearest neighbor classifier outperformed Bayesian classifier at measuring joint angles using accelerometers and gyroscopes for rehabilitation exercises, but their findings were based on only a small data set. Lum et al. [75] showed that K means clustering, K nearest neighbours, random forest, linear support vector machine and radial basis function support vector machine classifiers all worked better than a count ratio at measuring the functional use of an arm. An accelerometer on the wrist was used to collect their data. This demonstrates that machine learning models are capable of classifying human moment.

A physical activity questionnaire was compared to accelerometer data to see how accurate patients' perception of their physical activity were [41]. The study found that agreement between the questionnaire and the accelerometer data was low and that on average the participants did less physical activity than they reported in the questionnaire. They concluded that using accelerometers to measure physical activity would reduce bias and increase precision, which is plausible as it would remove the subjectivity in the assessments.

A few studies have used an IMU attached to both ankles to measure stride length. Stride length cannot be physically measured with an IMU system but can be estimated using calculations. One study [76] used a calculation developed by Rampp et al. [77]. The calculation works out a person's stride length using an accelerometer and gyroscope attached to the ankle. The calculation has an absolute error of about 6.26 cm. Ibrahim et al. [76] used the stride length and other spatio-temporal gait parameters to predict the fatigue of MS patients. The prediction was done using a random forest regression algorithm. Rudisch et al. [78] compared two IMU systems against an optical and three pressure-based systems for measuring stride length, cadence, heel strike and toe-off. The IMU systems measured stride length and cadence to a similar accuracy to the other systems.

In summary, Activity monitors can be versatile and adapted to the desired need. Terrain classification has been achieved, which is a measure defined in the K level definitions, but the method presented may not be suitable for lower limb prosthetic users. Machine learning models have been utilised with activity monitors to classify different measures.

1.3 Conclusions

Lower limb loss is becoming an increasing health problem worldwide, in part due to diabetes mellitus. Following lower limb loss, prostheses for TT and TF patients aim to aid with mobility. At present, K levels are assigned through clinical assessment that determine the prescription of prosthesis components, but issues remain due to the subjectivity of these K level assessments. There is also issues with the K level definitions, that are vague and require subjective interpretation. Satisfaction with lower limb prostheses remains low, and there is a need to improve matching components to patients. One solution may lie in the use of activity monitoring in the real world, to provide clinicians with objective every-day data on activities.

Review of the literature around activity monitoring revealed that a range of variables have been measured using a variety of different sensors or sensor combinations and classifiers to classify activity. However, no study has looked at utilizing activity monitoring in the context of K level assignment. Clinical practice remains without a solution to the shortfalls of subjective K level assignment. Hence there is a need for a sensor system and associated algorithms that can inform objectively on unsupervised activity of lower limb prosthesis users.

Design of such sensor-based system that provides objective, real-world activity data for lower limb prosthesis users involves a number of steps. To start with, it is important to understand the requirements that such a system needs to meet and which outcomes are clinically vital to clinicians. Once such understanding has been established, types of available sensors and classifiers can be reviewed, and those that have the potential to meet the requirements can be selected. Supervised real-world and in-lab data collection can then inform system design in terms of sensor placements and algorithm design. Once complete, it is then vital that the system and its output are tested in the context of unsupervised monitoring, and the clinical merit of the outcomes must be confirmed. Therefore, this PhD has the following Aim and Objectives:

1.3.1 Thesis aims and objectives

Aims:

This PhD aims to develop a clinically useful, sensor-based system that provides objective, real-world activity data to improve current K-level assessments for lower limb prosthetic users, specifically addressing current limitations in K-level classification, particularly at the K2-K3 boundary.

Objectives

- To investigate the clinical requirements for the objective system to aid in clinical decision making through interviews that explore clinicians' perceptions regarding shortfalls of current clinical activity assessments for K level assignments, and which objective measures they feel would improve their K level assignment. (Chapter 2)

- To review the literature around sensors and algorithms to inform system design. (Chapter 3)
- To design a system comprised of sensors and algorithms that output these measures, utilizing both, real-world and in-lab data collection. (Chapters 4 and 5)
- To evaluate the clinical benefit of the developed system and its outcome measures through clinician feedback on real lower limb prosthetic users' data, identifying areas for further improvement and development

Chapter 2

Investigation of clinical requirements for objective system to aid in clinical decision making through interviews

2.1 Background

In Chapter 1, the need for a sensor system that has the potential to objectively measure real-world activity in lower limb prosthesis users, to improve K level assignment, was established. Chapter 2 focuses on identifying the clinical requirements for a novel sensor-based system intended to enhance the accuracy of K-level assessments in lower limb prosthetic care. Before designing such a system, the clinical requirements need to be explored, to ensure that the system meets clinicians' needs. The **purpose** of this chapter hence was to conduct qualitative interviews with clinicians to gain an understanding of the clinical requirements for such an objective system, and thereby ensure it has the potential to aid in clinical decision making when assigning K levels for prescription of lower limb prosthetic components.

Qualitative research is employed to gather a holistic view of the phenomenon of interest in its context [79]. Glasziou and Chalmers estimate that 85% of health research funding is wasted but they suggest that most of this could be avoided with better designed research [80]. Slattery et al. argue that co-designed research would produce better designed research [81]. Co-designed research, sometimes called participatory research [82], is an approach where stakeholders for the research are engaged to create more meaningful and relevant research outcomes [81]. As the purpose of this PhD is to create a system to help clinicians in clinical decision-making for lower limb prosthetic component prescriptions, the end users of the system are clinicians who make clinical decisions for the prescription of lower limb prosthetic components. In consideration of this and to ensure the system is clinically relevant, qualitative research was conducted with clinicians, finding of which directed the development of the system in the subsequent chapters.

2.1.1 Aims of Chapter 2

- To gain an understanding of the shortfalls of current clinical activity assessments for K-level assignments.
- To determine the specific objective measures that clinicians would find most valuable for improving K-level assignment accuracy.
- To outline functional and practical specifications for a sensor-based system that aligns with clinical needs thereby enhancing decision-making for lower limb prosthetic prescriptions.

2.2 Methodology

Before developing an activity classification system, the clinical need(s) that the system is to meet and associated outcome measures of clinical relevance must be identified. For this, the views and experiences of prosthetic clinicians needed exploring. Some of the most used qualitative research methods to obtain these types of data are interviews, focus groups and questionnaires. To fully understand the clinically-relevant requirements, follow-up questions to responses are vital to clarify answers. Hence use of a questionnaire was not considered for this research [83]. Focus groups, however, have shown to give depth and insight into a participant's experiences and beliefs [84], and can also reveal a more nuanced perspective on a topic [85]. Interviews, on the other hand, can produce a bigger range of views and focus on individual opinions rather than a group consensus [83]. It is known that K level assessments vary between clinics; hence, it is important that we discover the individual personal views of the participating clinicians and not just a consensus to which they all agree. Logistically, interviews are also easier to organise. Considering this and the fact that there is only a small number of prosthetic clinicians in the country, it was decided that semi-structured interviews would be used to gain a thorough understanding of the clinical needs that a sensor system has to meet to aid in K level classification. Ethical approval was granted to record, transcribe and analyse interviews (Ethical approval numbers for University: 1710, Appendix D.1).

2.2.1 Participants

The participants that provided informed consent for the interviews were identified through the International Society of Prosthetics and Orthotics.

Participants met the following inclusion criteria: experience with activity level assessments in users of lower limb prostheses, able to provide informed consent, and able to do an interview over the phone/video call in the English language. Table 2.1 shows the professional background and experience of the participants in years.

Table 2.1 Participants professional background and experience

Participant	Background	Years experience	Type of clinic
P1	Prosthetist	27	Blatchford
P2	Consultant in rehabilitation medicine	52	NHS
P3	Prosthetist	16	Blatchford
P4	Amputee specialist physiotherapist	20	NHS
P5	Prosthetist	25	Blatchford
P6	Prosthetist	12	NHS

2.3 Data Collection: Interviews with clinicians

The aim of the interviews was to gather an understanding of how activity assessments are presently carried out at different clinics across the United Kingdom, explore the views of the clinicians regarding K levels and the associated assessments for K level assignment, investigate details of any past studies concerned with activity classification that the clinicians have been a part of and their views on these, and finally what data the clinicians would ideally want to help them in clinical decision making and their views on a sensor-based system that could potentially provide the needed data to assess patients in the real world.

To obtain the necessary information, an interview guide was developed, influenced by Kallio et at. [86]. The interview guide (Appendix C.1) was sent to all participants before the interviews to give participants an understanding of the type of questions that would be asked and why.

Due to Covid 19, all the interviews were conducted via video call, and audio was recorded using an external General Data Protection Regulation (GDPR) compliant recording device. The audio was then hand-transcribed, and any identifying details of the participants were removed from the transcripts. The interviews were conducted between May and July 2021. The interviews took no longer than 40 minutes to complete.

2.3.1 Data Analysis: thematic analysis of transcripts

The content of the transcripts was split into three main sections:

1. Current assessments
2. Past studies the participating clinician had been involved in
3. Views on a new sensor system that would provide a means to assess the activity of patients unsupervised in the real-world

These sections of the transcripts were analysed using thematic analysis based on a framework approach described by Braun and Clarke [87].

2.4 Results

2.4.1 Current assessments

The main themes highlighted about the current assessments were: 1) Assessment tests used; 2) Who is involved in the assessments; 3) Distinguishing between K2 and K3 patients; 4) Non-use of biomechanical data; 5) Timeframe for stabilising at a K Level after amputation; 6) Change in K level over time, 7) Effects of motivation on K level; 8) Problems with current K level assessments.

Assessment tests used

The participants were asked about the outcome measures and tests that they used, to assess a patient's activity level. The list below highlights all tests that were mentioned during the interviews and the number of participants that mentioned them.

- Get up and go test – 3
- Timed walk tests – 3
- SIGAM (Special Interest Group of Amputation Medicine) – 2
- Socket comfort score – 2

- AMPPRO (Amputee Mobility Predictor with Prosthesis) – 2
- Video assessment – 2
- Trips and falls questionnaire – 1
- Barthel Index – 1
- Time to recover test – 1
- 10-metre speed test – 1
- Walking around a stick test – 1
- L test – 1
- Pain score – 1
- Hospital anxiety and depression scale – 1
- LCI (Locomotor Capabilities Index) – 1
- Functional Assessment Measure (FIM+FAM) – 1

A key insight from this list is that a range of assessment methods are presently being used and that consistency is lacking across clinicians interviewed. Only 6 out of the 16 assessments were mentioned by more than one of the interviewees and only 2 mentioned by three separate interviewees. No assessment was used by more than half the interviewees.

Along with these tests, the participants talked about other techniques that they use during their assessments. Five of the participants said that what the patient tells them about their activities of daily living is a very important part of determining the patient's K level:

> *"It is mainly on what the patient is reporting"* (P3)
>
> *"I think it's a combination of us checking, their kind of rehab milestones so that they actually do those activities, plus also what they're subjectively telling us about their lives."* (P4)
>
> *"Having a verbal assessment dialogue with the patient to understand what they do in their own everyday life."* (P5)
>
> *"It's more asking the patients and actually just having a knowledge of your patients and observing how they walk."* (P6)

Participants 1 and 5 both said they asked their patients if they used walking aids, what terrain they traversed, if they can change their cadence, the work they are involved in, and if they participate in any sports. The other participants did not specifically say they asked their patients about these measures but did mention them as factors when deciding on a K level.

Participant 3 also said that when they are determining a patient's K level that *"It's not just about how far they can go for and how long, it's about stability and comfort and security in reducing risks of falls and allowing them to achieve their goals"*.

In summary, most of the assessments are based on a verbal dialogue with the patient and, although validated outcome measures are being used, they do not have a huge influence on the prescription. This makes the assessments mostly subjective which may add to the unreliability of the assessments. There is also a large variation of what different clinicians look for when assessing patients. This may cause different clinicians to assess patients at different levels and may, therefore, result in reliability being low between patients assessed at different centres.

Who is involved in the assessments.

From what the participants stated, the group of clinicians who are involved in K level assessments seems to vary between clinics, but it is usually a multidisciplinary team:

> *"It varies a lot, but usually a prosthetist and a physiotherapist are available all the time. So, they tend to be, if you want a constant in the team, whether you have a doctor or rehab engineer or any the other team members involved is going to vary a lot from centre to centre."* (P1)
>
> *"In conjunction with sort of a physio assessment."* (P5)

Whereas Participant 2 said it was just the prosthetists that use K levels:

> *"In practice, K levels were used by prosthetists. Not by clinicians, or not by physicians at all."* (P2)

In summary, a multidisciplinary team assesses the patient, but it might only be the prosthetist that routinely uses K levels when prescribing a limb for the patient.

Distinguishing between K2 and K3 patients

Distinguishing between K2 and K3 is the aim of the research that these interviews will influence. Four main differences between K2 and K3 patients were highlighted by the participants, these were change of cadence, terrain and use of aids:

Change of cadence

Change of cadence was mentioned by 3 of the participants as a key factor:

> *"If they can walk 10% faster 10% slower than their preferred walking speed, and do it comfortably."* (P1)

> *"They might kind of say, oh, like I was crossing the road, I had to like really speed up because there's a bus coming."* (P4)

> *"For me, personally, the big difference between a K2 and K3 is their ability to vary cadence."* (P6)

None of the participants distinguished if cadence is more of a factor for TF or TT patients, so it has to be assumed that it is considered for both.

Terrain

Three of the participants brought up the ability of the patient to traverse different terrain as a distinguishing factor:

> *"Getting them to tackle different terrains, so walking up down ramps, up down steps."* (P1)

> *"The difference between K2 and K3 is about the environment and the application."* (P4)

> *"They come back and they're like, Oh, yeah, like I took the dog for a walk in the woods."* (P4)

> *"K3 would be pretty much independent, able to walk in most conditions freely confidently, K2 be the ones that won't be able to probably walk as far, maybe needing some sort of aid outside for rougher conditions."* (P3)

Use of aids

As mentioned by Participant 3 above, the ability to walk without a walking aid was also stated by Participant 1 as a requirement for a K3:

> *"Walk safely without having to worry about or carry a walking aid, then it is a K3."* (P1)

Participants 3 and 5 talked about how K2 patients will not be able to walk as far as a K3 patient:

> *"Won't be able to walk as far"* (P3)

> *"K2 would be, wouldn't be able to go as far with activity. Maybe have to stop for more rests."* (P3)

> *"In a six-minute time walk test, they're more likely to be able to walk, we'd expect them to walk at least 300 metres, if we get somebody then who walks a lot less than that. And we thought they're on the boundary, we probably say, well, you are presenting to us more as a K2."* (P5)

In summary, being able to change cadence, traverse different terrain, walk without using a walking aid and walking without the need to rest were highlighted as the four main differences between a K2 and K3 patients. Being able to change cadence comfortably and to a degree of 10% was suggested. The types of terrains that were considered important were steps, slopes, curbs, and uneven ground and being comfortable at traversing them. The emphasis is to be able to walk safely without aids, so the stability of the patient is important.

Non-use of biomechanical data

All the participants were asked if they use or have used biomechanical data in assessments of patients. Biomechanical data were explained as objective data like gait symmetry, joint angles, or cadence. Two of the participants said that they or a member of their team has collected biomechanical data to help with patient assessments:

> *"With the microprocessor technology, we have got the ability to step count on the knees. So, we have switched them on."* (P3)

> *"We look at stride, we look at timing, we look at gait symmetry and we get*

a gait profile score. And we look at, we've for a while, we were doing energy expenditure as well with the O2 consumption, we're looking at knee flexion angles and things like that." (P6)

When asking which of these variables they thought was the most relevant, Participant 6 said:

"I think personally, I think the symmetry is the most important thing and making sure that we're getting a nice reciprocal, symmetrical gait pattern, and even weight bearing through both limbs." (P6)

But they also said:

"We don't regularly do it in our assessments." (P6)

This was echoed by the rest of the participants:

"We don't mainly because it's not routinely collected on patients." (P1)

"I don't think that is required routinely." (P2)

"We don't use it to direct our rehab and we don't use it routinely to allocate K levels." (P4)

"We don't do any, any measures." (P5)

A couple of the participants said that the reason why they don't collect biomechanical data is because they don't have access to it, with Participant 4 saying *"Nope. Mainly because we don't have access to it in the clinic"*.

In summary, biomechanical data are not commonly used to allocate or help allocate K levels. The inaccessibility of this type of data seems to be a common reason for it not being used. Step count, recorded by sensors inbuilt into microprocessor knees, is the most widely used data but not routinely used in assessments. Gait symmetry was highlighted as a variable of clinical interest.

Timeframe for stabilising at a K Level after amputation

When asked how long it would take a patient to settle into a K level after amputation, 3 of the participants agreed on six to twelve months:

> *"About the six months to a year point, after amputation."* (P1)

> *"Probably months six to twelve probably becomes established, some quicker depending on the level."* (P5)

> *"We would review a brand-new patient at one month, three months, six months and twelve months. And I'm hoping by the time we get to twelve months, we're reaching sort of peak of where the patient would be activity level-wise."* (P6)

Participant 4 gave a broader range:

> *"We probably would take about three months to rehab a transtibial, and about six months to rehab a transfemoral. But there will be patients that take, some patients are on our books for like a year and a half, because they take so much longer."* (P4)

Participant 1 elaborated on their answer and explained that in the first six months there is a lot more than just the physical element of amputation to deal with that can affect rehabilitation:

> *"I think in the first six months, there's so much information being bombarded at the amputee, they're having to come to terms with the amputation, and also the response of their family and friends to the amputation."* (P1)

In summary, it takes six months to a year after amputation for a patient to stabilise at a K level. There are patients who could reach it in three months and some that will take a year and a half, but participants overall agreed that most will take six months to a year. This implies that assessing a patients activity might not be necessary until six months post-amputation.

Change in K level over time

Most of the participants mentioned that K levels can change over time. Three said that patients can go up K levels:

> *"Yes, we do typically, yes to somebody who may be K2 initially, as you become more confident and following their physical and doing every bit more outdoor walking, we do sometimes see that they progress to a K3."* (P6)

> *"We've had people get remarried and things like that and find a new partner*

> *or engage in a hobby or, you know, get a new zest for life, then they tend to, they can go up."* (P3)

> *"Some people do change, some people go up the bands, and some people never ever achieve where you think."* (P1)

Participant 3 also echoed Participant 1's comments that K levels can go up and down:

> *"It does change, because if someone says has a stroke or their medical status changes, then obviously that's going to affect it."* (P3)

Participant 5 also talked about how health can affect the patient's mobility and therefore change their K level:

> *"Have a bad year or six months of health, then quite often we do see that there's a significant change in their prosthetic mobility status."* (P5)

They also said:

> *"You'd maybe get sort of 10% of that age group, maybe that change K class significantly."* (P5)

In summary, a patient can change their K level, even after they are established. Due to the age of most patients, they will gradually decline in their mobility and therefore their K level, but if a patient is given the right motivation, they can become more active and go up a K level.

Effects of motivation on K level

Participant 1 and 6 stated that the patient's motivation can affect their K level:

> *"I've had young, otherwise healthy, traumatic amputees, who you think this, there's no reason why this amputee, can't go on be a K 4, and they never do it because they don't, they're not motivated."* (P1)

> *"I think motivation and confidence are two of the big things that distinguish K2 and K3."* (P6)

Hence, motivation can be an important driver as to what K level a patient achieves during their rehabilitation, and as shown above, can also affect their K level as an established patient.

Problems with current K level assessments

All the participants were asked about their views regarding problems and issues with the current assessment methods. Three main areas were highlighted.

Defining the levels

The definition of the K levels was brought up as an issue by 3 of the participants:

> *"The definition of these K 1234 was so broad."* (P2)

> *"K2 K3 is like you're really tricky band to sort of understand the difference between them."* (P4)

> *"There isn't anything that would define, I've seen anyway. And to see that a K3 amputee of transfemoral or whatever level should be able to walk X number of distance at this speed, etc. So, we haven't had, been able to use anything to define, it's always been a little bit blurred."* (P5)

Subjective

Three of the participants said that the subjectiveness of the assessments can be an issue:

> *"What's the 10 for one patient might be another person's 5, you know, say in terms of like, socket comfort scores? I think, in prosthetics, there's a lot of things that is subjective, which is then very hard to compare patient groups against patient groups."* (P3)

> *"It's all very kind of, from our own experience, and practice that we're kind of making those judgments."* (P5)

> *"I think what I might see and judge to be a K2 or a K3 is not, or a colleague may judge as a K2 or a K3. So, I think it's very dependent on the clinician, and specifically what they determine a K2 or a K3 to be."* (P6)

Unreliable

Participant 1 talked about a previous experience where their team had classed a patient as K2 and not very mobile, but found when the patient wore an activity monitor that they were mobile and more in line with a K3 patient:

> *"No one who ever looked at him, from when we predicted to monitor and the consultant, the physical therapists, the clinician who looked after the*

> *full team, and myself, and the other prosthetists who were doing this study looking at the step counts, and cadence, because the monitor also recorded the cadence. And none of us honestly thought he was going to be doing it, we'd really thought it was gonna be a low user. Couple 1000 steps a day type person really, really surprised what he was doing."* (P1)

Participant 6 commented on how the physio's assessment can be different to the prosthetist's assessment of the same patient:

> *"If we compare them to our physio colleagues who use a different assessment tool, they typically score higher on the physio assessment tool, which is the AMPPRO than the observational score that we would assess."* (P6)

And when asked if there was a discrepancy in the assessments which one they would go with, they said *"I would go with my observational one."* (P6).

Patients not being knowledgeable of the assessments

Participants 1 and 6 mentioned that the patients are not very knowledgeable about the assessments and usually accept what the clinicians say:

> *"I think he just got on with it and thought this is the type of leg I should have, these people, they also know what they're doing, and no one had ever told him he could get something that was going to make things easier."* (P1)

> *"I don't actually think that patients know what K levels actually mean, or what an AMPRO score actually means."* (P6)

In summary, the definitions of the K levels and how to distinguish between them were highlighted as the two main issues, with some specific comments being made regarding differentiation between K2 and K3 in particular. Both may result in inconsistency in the allocation of K levels at different clinics. The subjectivity of the assessments, and how the assessments are being carried out was said to be unreliable by Participant 1 and 6 compared to validated activity assessment methods. The patients' limited knowledge of the assessments furthermore may contribute to inconsistent assessments, as they may not question any inconsistencies in prescriptions between patients.

2.4.2 Past studies the participating clinician had been involved in

The main themes highlighted about past studies conducted by the participants were: 1) Experience with activity monitors; 2) Positives of activity monitors and 3) Negatives of activity monitors.

Experience with activity monitors

All the participants were asked about their previous experience with activity monitors. Five of them had some experience of activity monitors attached to patients and four had used them as part of a study:

> *"We actually fitted an activity monitor to a group of patient's legs. And then we recorded what they did, it was using a system that in effect just printed out a diary at the end of it, told you on day one you walked a thousand steps day two you walked so many thousand so on so on."* (P1)

> *"We've done some studies with activity monitors. And they've been quite useful to have a look at step counts."* (P4)

> *"We did use them over 10 years ago, 15 years ago, there was quite a lot of work around using activity monitors."* (P5)

> *"We used to use the activity monitors the step counters and likes to calculate how many steps a person was doing over a period of time. So, I think they were called lam ones maybe or something like that. But they're miniature activity monitors that would fit in the prosthesis."* (P6)

Participant 1 also described the type of monitors that were used:

> *"They could collect data for up to a year, when we were trying to look at the long term, you should offer cause and effect, somebody could easily increase their activity level for a week. If they thought they were going to get better quality prosthetic care, or better type of limb out of it. But if you look over say a month you could see that."* (P1)

Participant 3 had not been part of any study but had a patient who used an activity monitor to monitor their rehabilitation:

> *"A recent quad patient, and she is monitoring her counted because she's a middle-aged lady. And she's recently lost both arms and legs. So, it has been good for her to measure. So, she's got a step counter, and she'd be measuring."* (P3)

Positives of activity monitors

A few positives from the participants were highlighted concerning use of activity monitors.

Change in K levels

As previously mentioned, Participant 1 elaborated about the patient who proved to be more active than they had been clinically-assessed, and therefore the activity data resulted in a change of his K level classification and the prothesis he used:

> *"We'd saw him for years. And he used to come in and tell us that he took his dog out for walk, but his leg was immaculately clean and think, never doing anything. He pushes his wheelchair into the limb centre, he'd be treated, and he'd pushes his wheelchair out the back but when we stuck an activity monitor, when he came back after a while he was doing 6 / 7000 steps a day. When we looked at it, he was taking his dog for a walk three four times a day and we had him on like a SACH foot. I don't know if you know much about prosthetics but SACH feet are really basic, fairly naff foot. It's not much good for anything. And then we were like oh we better upgrade him fast because he's actually doing something. We had him down as a like as a K2 user, somebody who really didn't do much at all."* (P1)

This indicates that activity monitors can be used to adjust a patient's prescription to ensure the patient has a prosthesis that is more suitable for their activities. Participant 1 also clarified that they did not use the data to limit any patient's prosthetic care:

> *"We didn't really limit people's prosthetic care as a result of it or, did change the classifications of a few people who moved up, because we saw they were actually doing an awful lot more than we believed."* (P1)

From this it can be obtained that clinicians are willing to use activity data to increase a patients K level if the data suggests their K level is too low but not reduce their K level or prescription if they have lower activity than perceived. An explanation for this could be that higher than perceived activity levels proves the patient is capable of higher activity levels but lower recorded activity level do not necessarily prove the patient is not able to produce higher activity levels, as other factors such as the weather or an acute health condition could have affected their activity levels during the recording period.

Motivation

How the data from the activity monitors motivates the patients to be more active was a positive mentioned by Participants 1 and 3:

> *"What we found it was because the amputees come back to the rehabilitation centre so often, they get to know a lot people and also if you have two amputees who had their amputation at the same time, they tend to keep in touch with each other years afterwards. So, then they were comparing the results. So, there was quite a good buy into it."* (P1)

> *"Now she's managing to do 5000 steps a day, from not being able to walk. So, she's found in that good for her. So, she's using that technology herself to, to monitor herself and then say yes, I've achieved in 5000 steps a day. So, I suppose it's spurring her on. But not everybody is as motivated as that. But it could be used as a motivating factor."* (P3)

New data

Participant 1 talked about how the data the activity monitors produce was something the clinicians hadn't seen before, and which excited them:

> *"Everyone who got involved in the project and helped out in it at one part or another, whether it be the patients, the physiotherapist, the prosthetist, they were all really quite excited about it. Suddenly the doctor rehabilitation consultant, at the time got really excited about it he thought was an amazing thing, because he'd never saw that kind of information about patients."* (P1)

Patients' acceptance

Participant 4 mentioned how accepting the patients were to participating in the studies and having the activity monitors attached to them:

> *"They've been quite acceptable to patients, especially if they're attached to the device, rather than to the patient themselves."* (P4)

> *"I think they're also quite acceptable for patients to apply to themselves, which is quite useful then if the study says you don't have to recall the patient back to the centre to put the Activity Monitor on, you can just send the Activity Monitor out to the patient, and they can put it on their own device."* (P4)

> *"We have lots of patients here who are usually quite happy for their activity to be monitored."* (P6)

In summary, activity monitor data can be used to change a patient's K level, providing clinicians with valuable, previously unavailable insights into patient activity. The data has also been shown to motivate patients to be more active. Clinicians have been excited by the data because it is new information about their patients that they do not usually get to see. Additionally, the monitors are generally well-accepted by patients, especially when integrated into the prosthesis rather than attached directly to the body.

Negatives of activity monitors

A few negative aspects of the participant's experience of activity monitors were brought up during the interviews.

Patient's refusal

Participant 1 talked about two patients that had not wanted to be a part of the study:

> *"We only had a couple of people who said that they weren't interested in getting involved in it. And usually those were people who are very worried about the cosmesis, because we're putting it inside the legs, so you can't tamper, tamper with it."* (P1)

Participant 4 also mentioned that where the monitor is attached can affect the acceptance of the patients to having them attached:

> *"Then sometimes have to attach it to their body. And that's not as useful or not useful, it's not as acceptable to patients."* (P4)

Participant 6 also stated that some patients don't allow for the activity monitors built into MPKs to be switched on because they are worried about losing benefits:

> *"We also have patients who are quite reluctant, because they're worried that maybe their benefits may be affected, because they're actually better than they're maybe making out to be."* (P6)

Limitations of activity monitors

The limitation of some activity monitors to differentiate between sitting and standing and if the patient is not wearing the prosthesis was commented on by Participant 4:

> *"The trans-tibial, that are trickier because when you put it on, if you're going to put it on their prosthesis, then it does, it can't tell the difference, then between whether they're, they're stood or they're sitting."* (P4)

> *"The other downside actually with, with activity monitors is with prosthetics is we have to tell the patients to lie the prosthesis down because otherwise the Activity Monitor thinks the patient stood up all night."* (P4)

Feedback

Participant 4 mentioned that the data that were produced by the activity monitors was too complicated and didn't just present the important information they are looking for:

> *"When you look at it, it's a bit like, oh, god, what exactly is it that I'm looking at? And so, I suppose again, it comes back down to that thing of like, what's the most important data to present? How do we know what the most important data is? And how do we get rid of like all that noise that actually just confuses things and make sure that we're presenting stuff that's really meaningful."* (P4)

They also criticised activity monitors for not producing real time data:

> *"I think the big downside is that they don't give you any like real time feedback."* (P4)

Cost

The cost of the monitors was a restriction that stopped Participant 1's clinic carrying on with using activity monitors:

> *"We've got about 1700 amputees. And if you're going to need If, even say only half those amputees have two limbs, you're still talking about 2000 plus devices. And it just worked out the cost of it was, people were saying, Wait a minute, that's a big chunk of money. It's like a month's budget."* (P1)

In summary, participating clinicians mentioned that a few patients refused to participate in the studies; the main reasons for not participating were having the activity monitor attached to their body, damaging the cosmesis of the prosthesis, and fear the data could be used to change their benefits. The limitation of the activity monitor not to be able to distinguish between sitting and standing and not being able to identify whether or not the patient was wearing the prosthesis was another negative. The data presented to the clinicians was at times confusing and not always the type of data that is important to them. Participant 4 also preferred real-time data feedback. Finally, the cost of the devices was highlighted as a reason why participants did not continue to use activity monitors.

2.4.3 Perspectives on a real-world activity monitoring system

During the interviews the participants were asked about their views regarding the development of a new sensor-based system that would be attached to the prosthesis to monitor the patient outside of the clinic. The main themes that were highlighted were: 1) Desired real-world data; 2) Feedback of the system; and 3) Further considerations.

Desired real-world data

When asked what their desired data would be that the system should measure, the main variables that were mentioned were: step count, cadence, the type of terrain they walk over, wear time and distance covered:

> *"How many steps they're taking, but also cadence. And it has to be a long-term thing."* (P1)

> *"Distance covered, I guess. So, step, I suppose that time step length wouldn't it? Yeah."* (P3)

> *"I suppose it would be useful if you were able to detect information about the terrain that they're walking on, whether you were able to detect the speed that they were walking, and whether they're able to change that speed, because obviously, that's an important thing about the K2 K3 is about that variable cadence. I suppose the other thing would be things like, you know, like, if they like stepping, like going up and down steps and curbs."* (P4)

> *"I think I would probably think that looking at the overall distance that somebody has, has walked would be interesting. And the various, and the speeds and the variable variation of their speed over that time would be useful"* they also said *"terrain. Yes, that would be a good one as well to see because that would link in with their activity coding because it's showing that they are navigating over obstacles."* (P5)

> *"To get a good idea of what a patient is doing outdoors and the types*

> *of terrain they're walking on. I think that would be good. And how long and are they they're actual wearing and using their limb."* (P6)

Feedback of the system

Three participants discussed how the system would deliver feedback on the data. Participant 1 talked about how the data could be used for gait training during rehabilitation:

> *"Some devices on new amputees that gave me step counts and cadence. And then I showed the early input on gait training to give education actually improved overall the patient's mobility periods."* (P1)

Participant 2 also expressed a similar view that it could be useful for gait training:

> *"If you think in terms of a subsection of improving the person's gait, in an appropriate prosthesis, in other words, the prosthetic element part that yes, it will be, it could be helpful for the prosthetist."* (P2)

In addition, Participant 4 talked about making sure the data is clear and easy to understand:

> *"How usable is it like, how does it, does it collect something meaningful? And, and then when you're looking at it on the screen? Like, is it in a format that's really easy to understand? Because you think, most clinicians are not like academics, they need things pretty straightforward, pretty simple and straight to the point."* (P4)

Furthermore, the usefulness of real-time feedback was discussed by three of the participants. Participants 1 and 4 said that real-time feedback could be useful to help motivate patients to be more active:

> *"If they can see maybe it's tagged on the phone as an app or whatever, then I think you would end up with a really good compliance for the users. Because amputees themselves, because they would like to do it, see that kind of information. And I think it's, it's one of those things, because then if they can, then come back into the next appointment, say, here you go I've have managed to do 7000 steps a day or 10,000 steps a day whatever it is, then you're good."* (P1)

> *"Having data in real time is really useful. And again, it's like also, we've had lots of reading activity monitors in different studies, there's also lots of lots of kind of feedback that patients quite like to have that information as well. And that can be quite motivating for them to do more."* (P4)

Participant 6 was a little more hesitant about its usefulness, saying *"It may be useful for certain instances, but usually it would be when you're reviewing the patient."* (P6).

In summary, cadence, terrain, distance and step count were mentioned by participating clinicians as desirable data to support their assessment of K levels, with wear time also suggested by one participant. A couple of the participants thought it would be ideal if a new system could assist with gait training. It was also highlighted that the feedback must be simple and straightforward so that it is easy for clinicians to understand. Finally, real time feedback was discussed with some participants suggesting that it could be useful to motivate the patients to be more active, but it probably would not affect the clinicians' assessments.

Further considerations

A few further points were raised by the participants that should be considered when developing the new system.

Consent

The consent of the patients was raised by a few of the participants as something that will need to be considered:

> *"There was a, a question raised about the ethics and consent."* (P2)
>
> *"I mean, they would have to have their consent to do it."* (P5)
>
> *"I guess within the NHS there's a big thing about consent and patient consent and whether they consent to the Activity Monitor being placed there."* (P6)

Weight and size

Two of the participants mentioned that the weight and size of the system will need to be thought about and maybe also waterproofness:

> *"I think it would be useful as long as it wasn't heavy. And then didn't impact with anything else."* (P3)
>
> *"I guess it would depend on the size of the monitor and how it may attach. We have patients that have legs that they use for walking but for swimming, and it's going to be waterproof."* (P6)

Data comparison

Participants 3 and 5 brought up the issue of what to do with the data, i.e. what to compare the data against and what would the data mean:

> *"Yeah, I think it would be probably useful if we did it on everybody. But there's not that much. As when we did the trips and falls, there's not that much detail out there that shows you what an average amputee does for this level. So, if there's, at the moment there's not much data out there that we could compare it against."* (P3)

> *"I'd like to see information that I could look up to say, I've done these measures, how do they compare to what to K2 do, which category do they fit into?"* (P5)

Hence, ensuring that patients' consent is obtained will be vital. The weight and size of the device were also highlighted as critical to be considered, and the device should not change the performance of the prosthesis. What the data means will be the most challenging consideration. The data will not be useful to clinicians if there is no clarity as to what results each K level should achieve, hence baselines for each level will need to be determined.

2.5 Discussion and Conclusions

Study 1 identified through interviewing clinicians that the main issues of reliability in the current activity level assessments come from the subjectivity and variability of the assessments. Most of the assessments are dialogue-based where different clinicians ask different questions. It was shown by Limb et al. [41] and Balkman et al. [42] that dialogue-based questionnaire assessments do not give an accurate assessment of a patient's physical activity. The objective outcome measures that are used vary from clinic to clinic, and although most of the ones used are validated, the observational subjective assessment takes priority, as stated by Participant 6. There is also some variance in how the clinicians interpret the K levels and especially the boundary between K2 and K3. The ambiguity of the levels and the boundaries was brought up by a few of the participants and is something that needs to be addressed, but that will be more related to how the different prosthesis components benefit patients with different activity levels.

A few variables were highlighted as the main differences between a K2 and K3 patient. The ability to change cadence, traverse different terrain, walking without an aid and covering a greater distance were the main four. These were, as expected, echoed with the desired data that the participants would want from a new sensor-based system, but with the inclusion of step count and possibly wear time. From these measures only cadence is specifically stated in the current K level definitions, Table 1.1. This suggests that the current K levels definitions might not meet the clinical needs for prescribing prosthetic components and therefore an update or replacement to K levels could have clinical benefit. Changes in cadence and step count have been measured accurately with an accelerometer or IMU in several studies [65][66][67]. Wear time has been measured using a proximity sensor or a force transducer to a high accuracy [71][65]. Distance covered can be accurately measured using a combination of GPS and an accelerometer [69] or from step count and an estimated stride length [77][76][78], whereas Hoeger et al. [88] found that step length changed at different cadences. Bassett et al [89] found that step count alone has a high correlation to physical activity and health. From this, it might be more accurate and beneficial to record step count and cadence than distance covered. No prosthesis-mounted sensor-based system has previously measured a patient's ability to traverse different terrains and walking without an aid. Camargo et al. [73] used IMUs and goniometers to measure ramp incline and step height, but a goniometer would not be effective on a prosthesis and the exact height and incline would not be necessary to assess if a patient is traversing steps or a ramp. To fully understand what sensors are required to assess if a patient is traversing a certain terrain or is walking without a walking aid, a greater understanding of the characteristics of these types of locomotion is needed.

K levels seem to settle within six months to a year after amputation, hence a system to assess the patient's activity may only become useful after this period. But there was some desire for the system to be able to help gait training, either by just counting steps or trying to incorporate a more complex system that measures a variable like gait symmetry. K levels may change over time, with found or lost motivation and changes in health acting as drivers to changes in K level. If a sensor-based system could give feedback to the patients, whether that is in real time or not, it could help motivate the patients to be more active, which could be beneficial to their health. The system feedback to the clinicians must be clear and easy to understand, as clinicians do not have the training or time to understand complicated data. But from previous trials the participants have been a part of, the clinicians and patients have been excited to see the data, hence there is a desire to engage with it. Two of the participants also asked for comparative or normative data to help assign a patient into a K level by comparing their activity levels against typical patients for each K level. For this a large dataset would be needed, and as such could not be achieved within the scope of this project.

Conclusion:

Current K-level assessments are subjective, with vague definitions and inconsistent application across clinics, leading to variability and reduced reliability in prosthetic prescriptions.

Clinicians identified four critical objective measures for improving K-level assessments:

- Cadence
- Use of walking aids
- Terrain traversed
- Step count

A sensor-based system that captures these objective measures could enhance the accuracy and consistency of clinical decision-making for lower limb prosthetics.
For the system to be helpful in practice, the data must be presented in a simple, user-friendly format that clinicians can easily interpret and apply.
This chapter established the clinical requirements for the objective system to aid in clinical decision. The next chapter (Chapter 3) reviews the literature around sensors and algorithms to inform the design of an effective real-world monitoring system.

Chapter 3

Review of sensors and algorithms for activity classification to inform system design.

3.1 Background

In Chapter 2, the clinical requirements for the objective system to aid in clinical decision making were established through interviews that explored clinicians' perceptions regarding shortfalls of current clinical activity assessments for K level assignments, and which objective measures they feel could improve their K level assignment. Critically, the interviews demonstrated that the measures such a system needs to provide include steps, cadence, terrain and walking aid use.

As the next step towards developing a sensor system that meets clinicians' needs, the **purpose** of this chapter was to evaluate existing sensor technologies and machine learning algorithms that could support accurate, real-world activity monitoring. By examining prior research on sensors and algorithms, to inform the design of the sensor-based system for real-world data collection reported on in the subsequent chapters.

3.1.1 Aims of Chapter 3

- Review machine learning algorithms and sensor systems used for unsupervised activity classification in real-world environments, which will be needed for terrain and walking aid recognition.

Review of sensors and algorithms for activity classification to inform system design.

- Review specific sensor technologies and their specifications for application in unsupervised activity classification systems to inform the selection of both sensors and algorithms to be used in subsequent chapters.

However, before a thorough review is presented on algorithms and sensors for activity monitoring in general, this chapter briefly discusses the only two papers on recognition of walking aid use specifically. Their discussion is then followed by the review of algorithms and sensors that are most relevant for developing a robust activity classification system that has the potential to meet the clinical need.

3.2 Walking aid recognition from wearable technologies

The only two studies found which investigated walking aid recognition using wearable sensors were by Antos and Moder [90][91]. Both achieved high accuracies using a wrist worn accelerometer, but both of these studies were conducted in limited indoor environments so the performance on a variety of terrain was not investigated. The system that this research is developing has to limit any effects on the prosthesis user's life. Asking a user to wear a wrist worn device for a long period might reduce the acceptability of the system.

Studies have found differences in lower limb kinematics during strides with and without a walking aid on flat ground [92][93][94] and stair climbing [95]. Wearable sensors on the lower limbs are able to measure the kinematics. Hence, the use of wearables on the lower limbs should be able to detect the differences between strides with and without a walking aid, and this insight will be taken forward into the subsequent work.

Attaching an accelerometer or IMU to the prosthesis users walking aid could be an accurate measure of how much the users uses that specific walking aid. The drawback of this method is that all walking aids the patient uses will have to have a sensor attached and it would not be detected if the patient used a hand rail or other supporting structure as an aid. for these reasons this method was not considered for this research.

3.3 Review of machine learning algorithms for unsupervised activity classification in the real world.

Overview

A systematic review concerned with locomotion recognition in users of assistive devices has previously been conducted [96]. The aim of the review was to examine the use of machine learning algorithms for locomotion recognition for assistive devices and identify areas for future research. The key area identified by the review was the current lack of studies that incorporated real-world environments. The review included studies that predominantly looked at locomotion recognition for activity predictions for microprocessor-controlled prostheses. Fifty-eight articles were reviewed that were published between January 2000 and July 2020. The review of these articles informed this chapter of the PhD and dictated the subsequent research. Since the review did not capture papers published after July 2020, an updated literature search was conducted for papers published between 2020 and 2024 for the purpose of this thesis. The same search strategy and eligibility criteria were used, as had been used in the previous review. The initial search found 1039 papers. After removing duplicates and screening, 75 papers were passed forward for full text review. Out of these, 42 were deemed to not meet the inclusion criteria, hence 33 additional papers (published since 2020) were freshly reviewed in this review chapter. In the following sections, details are provided on the search term and databases used, the associated output and its quality. This is then followed by a thorough discussion of relevant content of the original review paper and also the papers identified through the updated literature search.

3.3.1 Methods

Search strategy used to update the literature search.

In PubMed, the following search term was used:

((((exoskeleton OR exoskeleton robot OR powered exoskeleton OR powered lower limb exoskeleton OR wearable exoskeleton OR lower limb exoskeleton OR lower limb prostheses OR lower limb prosthesis OR transfemoral amputee OR powered prosthesis OR above-knee amputation OR powered lower limb prosthesis control OR powered above-knee prosthesis OR transtibial amputation OR prosthesis use OR powered prosthesis leg OR amputees OR orthotics OR orthoses OR orthoses OR orthotics)) AND (intent recognition OR locomotion mode classification OR terrain recognition system OR user intent recognition OR locomotion mode recognition OR pattern recognition OR user-independent intent

recognition OR terrain recognition OR terrain-adaptive system OR adaptive pattern classifier OR human motion intent)))

Regarding additional filtering, the following article types were selected:

Case reports, classical article, clinical study, clinical trial, comparative study, controlled clinical trial, evaluation studies, journal article, letter, multicenter study, pragmatic clinical trial, randomized controlled trial, patents and conference papers, and of these only studies on humans were included.

In Web of Science, the following search string was used:

(TS = ((Exoskeleton OR Exoskeleton robot OR Powered exoskeleton OR Powered lower limb exoskele-ton OR wearable exoskeleton OR lower limb exoskeleton OR Lower limb prostheses OR lower limb prosthesis OR Transfemoral amputee OR powered prosthesis OR Above-knee amputation OR Powered Lower Limb Prosthesis Control OR powered above-knee prosthesis OR transtibial amputation OR pros-thesis use OR powered prosthesis leg OR amputees OR Orthotics OR orthoses OR orthosis OR orthotics) AND (Intent Recognition OR Locomotion Mode Classification OR Terrain recognition system OR user intent recognition OR Locomotion mode recognition OR Pattern Recognition OR User-Independent In-tent Recognition OR terrain recognition OR terrain-adaptive system OR adaptive pattern classifier OR human motion intent))) AND LANGUAGE: (English) AND DOCUMENT TYPES: (Article OR Data Pa-per OR Letter OR Proceedings Paper)

The search results for papers published since 2020 are summarised in an adaptation of the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart shown in Figure 3.1.

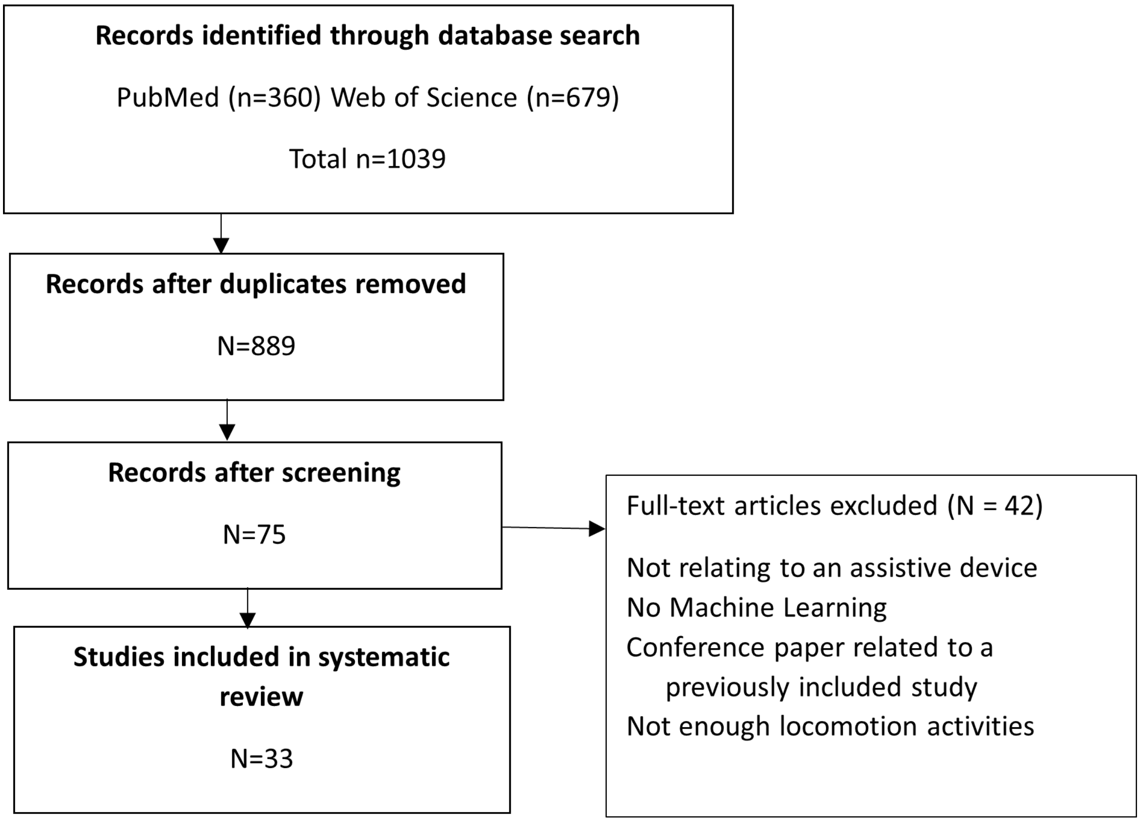

Fig. 3.1 Review search flowchart

Quality assessment of papers identified through the updated literature search. A quality score for each study was determined using the same Modified QualSyst Tool as had been used by Labarrière et al. [96], see below and also appendix G.1 for further details. Table 3.1 details the quality score, participant group and assistive device for each identified study. The mean quality score for the studies was 66%. In comparison, the mean quality score from the original review was 68%. This demonstration that the quality of papers included in the updated review are similar to the quality of the original review.

Quality score criteria

- Criteria 1: Question/Objective sufficiently described
- Criteria 2: Study design evident and appropriate
- Criteria 3: Subject characteristics sufficiently described and representative
- Criteria 4: Experimental protocol sufficiently described
- Criteria 5: Critical Timing Provided

- Criteria 6: Filtering method sufficiently described
- Criteria 7: Window length clearly mentioned
- Criteria 8: Input features clearly mentioned
- Criteria 9: Machine Learning algorithm clearly mentioned
- Criteria 10: Evaluation method of the machine learning algorithm clearly mentioned
- Criteria 11: Results reported with enough detail
- Criteria 12: Conclusions supported by the results

Each criteria was evaluated with a score between 0 and 2: 2 indicates "yes", 1 indicates "partial" and 0 indicates "no". Additionally, prior to assessing the quality of the studies, the following guidelines were created to ensure consistency in ratings.

The quality score for each paper was calculated as the sum of the scores for each criteria divided by the maximum possible score.

Table 3.1 Quality assessment and recruited volunteers in the included studies. TT=transtibial, TF=Transfemoral, NS=not stated, n/a=not applicable

Article	Quality score	Groups	Locomotive assistive device
Papapicco et al. [97]	75.00%	10 healthy	Prosthetic
Wang et al. [98]	50.00%	7 healthy, 1 TT	Exoskeleton
Lu et al. [99]	54.20%	10 healthy	Exoskeleton
Khodabandelou et al. [100]	83.30%	10 healthy	Exoskeleton
Li et al. [101]	87.50%	10 healthy	Exoskeleton
Gao et al. [102]	58.30%	10 healthy	Exoskeleton
Chen et al. [103]	62.50%	1 Healthy	Exoskeleton
Zhu et al. [104]	62.50%	7 healthy	Exoskeleton
Tang et al. [105]	66.70%	15 healthy	Exoskeleton
S. Gao et al. [106]	83.30%	10 healthy	Exoskeleton
Zheng et al. [107]	66.70%	10 Healthy, 1 TT, 1 TF	Prosthetic

Yu et al. [108]	62.50%	3 healthy	Exoskeleton
Chauhan et al. [109]	83.30%	30 healthy	n/a
Vu et al. [110]	45.80%	4 healthy	Prosthetic
Sun & Li [111]	37.50%	NS	Exoskeleton
Madaoui et al. [112]	83.30%	4 healthy	Prosthetic
Han et al. [113]	54.20%	6 healthy	n/a
Oh & Hong [114]	50.00%	1 healthy	Exoskeleton
Luo et al. [115]	50.00%	10 healthy, 2 hip amputees	Prosthetic
Zhong et al. [116]	62.50%	7 healthy, 1 TT	Prosthetic
Gonzales-Huisa et al. [117]	91.70%	24 healthy, 5 TT	Prosthetic
Yin et al. [118]	41.70%	22 healthy, 1TF	Prosthetic
Qi et al. [119]	62.50%	5 healthy	Exoskeleton
Marcos Mazon et al. [120]	79.70%	10 healthy, 1 TF	Prosthetic
Bruinsma & Carloni [121]	83.30%	1 TF	Prosthetic
F. Gao et al. [122]	54.20%	3 healthy, 3 TT	Prosthetic
Shin et al. [123]	66.70%	4 healthy	Exoskeleton
Son & Kang [124]	79.20%	500 healthy	Exoskeleton
Liu et al.[125]	58.30%	8 healthy	Prosthetic
Haque et al. [126]	54.20%	2 Healthy	Exoskeleton
Liu & Wang [127]	75.00%	5 healthy, 1 stroke	Exoskeleton
Guo et al. [128]	70.80%	10 Healthy	Exoskeleton
Qian et al. [129]	83.30%	10 healthy	Exoskeleton

The following sections discuss the content of the original review from 2020 as well as that of the papers identified from 2020 onwards under relevant headings.

3.3.2 Results

Assistive devices

The original review from Labarrière et al., [96] included 50 prosthetic-limb-focused articles, 6 about exoskeletons and 2 on orthoses. For the review update, assistive devices were categorised as prosthetics (N=12) and exoskeletons (N=19). Two studies used healthy participants and said the technology could be used for any lower limb assistive device [109][113].

Participants

For the Labarrière et al. [96] review, the mean participant number was 5.5. All of the exoskeleton and orthoses studies only used healthy participants as did 3 prosthesis articles. Out of the other 47 prosthetic articles, 32 only used prosthesis users and 15 used prosthesis and healthy participants. The average population of prosthesis user participants was 3.9. The largest prosthesis user participant number was 9 [130]. Twelve of the prosthetic studies only recruited one prosthetic user. Thirty-two of the prosthetic studies focused on transtibial prosthetic users and 18 for transfemoral prosthetic users. No study incorporated both transtibial and transfemoral prosthetic users.

From the review update, most studies used healthy participants, with only 10 studies using target populations. The mean size of the studies was 24.6 participants, but one study had a much larger population of 500 participants [124]. With this study excluded, the mean reduced to 9.2 participants per study. Out of the studies that had the target population as participants, the mean number of these participants was 1.8, the largest being 5. Only one study solely used the target population, but it had only one participant [121]. Out of the prosthetic studies, five studies used transtibial prosthetic users [98][107][116][117][122], 4 used transfemoral prosthetic users [107][118][120][121], and one used transpelvic amputees [115]. Only 4 studies recruited more than one prosthetic user, with one study having 5 transtibial [117], one having 3 transtibial [122], one having 2 transpelvic [115] and one study having a transtibial and a transfemoral prosthesis user [107].

The majority of studies used healthy participants rather than the target population. This raises concerns about the generalisability of the results, especially given the variability in gait patterns between healthy individuals and prosthesis users. Studies involving target populations used very small samples, questioning the validity of the accuracies for real-world activity classification in the lower limb prosthetic user population. This is a gap in the research that previous literature has not covered.

Movement tasks and terrain in activity classification.

For the Labarrière et al. [96] review, 43 studies incorporated flat terrain, stair use and ramp walking, with 39 of these focusing exclusively on these three activities and no other activity. Of the other studies, 13 involved flat terrain and stair use but did not investigate ramp walking, 6 of these only investigated flat terrain and stair use and no other activity. Obstacle clearance was incorporated in 6 studies and banked walking in one. No studies investigated any type of uneven or unstable terrain.

For the review update, Table 3.2 details the locomotive activities, sensors, data processing, algorithms and accuracies for each study identified. All the studies used flat terrain and only one study did not include stair ascending and descending [106]. Twenty-four studies used ramp ascend and descend. Uneven terrain was incorporated into one study [121], three studies used grass [98][116][109], one used cobblestones [109] and the same study used banked terrain. Twenty studies just looked at flat ground, stairs and ramps. All 9 of the studies that recruited prosthetic users collected data on flat and stairs with 2 also including grass [98][116] and the other 7 including ramps, with one of these also investigating uneven terrain [121].

The majority of studies reviewed focused on a limited range of locomotive activities, primarily flat terrain, stair use, and ramp walking. Very few examined more challenging terrains, only four out of the 91 total articles incorporated a type of uneven terrain, for example cobblestones or grass, and none of the studies investigated unstable terrains such as gravel or sand. This shows a gap in the research, particularly in relation to real-world environments where prosthetic users encounter various terrains beyond flat surfaces and standard inclines.

Table 3.2 Details of studies identified through review update. F=flat terrain, S=stair use, R=ramp walking, G=grass walking, CS=cobblestones, U=uneven terrain, B=banked terrain, SVM=Support vector machine, CNN= Convolutional Neural Networks, LSTM= Long short term memory, DBN= Dynamic Bayesian Network, QDA= Quadratic Discriminant Analysis, ANN=Artificial Neural Networks , DT=Decision tree, LDA= Linear Discriminant Analysis, NNR=neural network regression, BNN=Bayesian neural networks , GMM=Gaussian mixture model, HSVM=Hierarchical support vector machine, NS=Not stated, MAE=mean absolute error.

Article	Locomotive activity	Sensor	Online / offline	Recognition / prediction	Windowing Type	window Size	Sample rate	Algorithms	Accuracy (%)
Papapicco et al. [97]	F S	2 IMU, Pressure	Offline	Prediction	Stride	100 samples	100Hz	SVM	95.6
Wang et al. [98]	F G S	1 IMU, Camera	Offline	Prediction	Sliding	60 samples	NS	CNN, LSTM, ResNet, ResNet-Att	99
Lu et al. [99]	F S R	5 IMU	Offline	Prediction	Sliding	300ms	20Hz	SVM, LDA, QDA, ANN	95
Khodabandelou et al. [100]	F S R	2 IMU, Knee angle, Hip angles	Offline	Prediction	NS	3s	100Hz	CNN	99.78
Li et al. [101]	F S R	5 IMU, 14 EMG	Offline	Prediction	From gait event	300ms	500Hz	DBN NN	97.64
Gao et al. [102]	F S R	3 IMU, Force	Online	Recognition	From point	190ms	100Hz	LSTM	98.81
Chen et al. [103]	F S	8 IMU	Offline	Recognition	Sliding	128 samples	200Hz	LSTM	97.78
Zhu et al. [104]	F S R	4 IMU, 2 load cell	Online	Recognition	Sliding	100 samples	50Hz	CNN	97.64
Tang et al. [105]	F S R	3 IMU, 2 pressure	Offline	Recognition	Sliding	8 samples	100Hz	LSTM	97.93
S. Gao et al. [106]	F R	3 EMG, 2 Pressure	Offline	Recognition	Stride	NS	2000Hz	SVM	96.8
Zheng et al. [107]	F S R	2 IMU, Pressure	Offline	Recognition	Sliding	100ms	100Hz	fuzzy clasifier and dynamic time warping template	95.38
Yu et al. [108]	F S	7 IMU	Offline	Recognition	Sliding	150ms	200Hz	ANN	99.55

Chauhan et al. [109]	F S R CS G B	6 IMU	Offline	Recognition	NS	8s	100Hz	CNN+SVM	98.2
Vu et al. [110]	F S	1 IMU	Offline	Recognition	Sliding	50ms	1MHz	CNN, LSTM	99.6
Sun & Li [111]	F S R	1 IMU	Offline	Recognition	NS	NS	NS	CNN	98.6
Madaoui et al. [112]	F S R	1 IMU	Offline	Recognition	Strides	260 samples	200Hz	ANN	97.3
Han et al. [113]	F S R	1 IMU	Offline	Recognition	NS	200ms	100Hz	DT	96.71
Oh & Hong [114]	F S	3 IMU	Offline	Prediction	NS	4ms	NS	ANN	99
Luo et al. [115]	F S R	1 IMU, lazer range sensors	Online	Recognition	Stride	NS	50Hz	NS	98.67
Zhong et al. [116]	F G S	1 IMU, Camera, GPS	Online	Recognition	NS	500ms	10Hz	BNN	93
Gonzales-Huisa et al. [117]	F S R	3 IMU, 4 EMG	Offline	Recognition	Sliding	80ms	148Hz	SVM, LSTM	95.46
Yin et al. [118]	F S R	1 IMU, Camera	Offline	Prediction	Individual points	NS	NS	CNN	96
Qi et al. [119]	F S R	2 IMU, Pressure	Offline	Recognition	Stride	NS	100Hz	HSVM	97.106
Marcos Mazon et al. [120]	F S R	2 IMU	Offline	Prediction	Stride	500ms	500Hz	LSTM	95
Bruinsma & Carloni [121]	F S R U	2 IMU	Offline	Recognition	Sliding	30ms	1000Hz	CNN	93.06
F. Gao et al. [122]	F S R	1 IMU	Offline	Recognition	Individual points	NS	200Hz	Threshold based	98.5
Shin et al. [123]	F S R	4 IMU	Offline	Recognition	Individual points	NS	100Hz	GMM	99.33

Son & Kang [124]	F S R	1 IMU, 2 hip angle, 8 EMG	Offline	Recognition	Sliding	1.76s	100Hz	CNN	96.17
Liu et al. [125]	F S R	2 IMU, Pressure	Offline	Recognition	NS	50 samples	100Hz	ANN	99.16
Haque et al. [126]	F S	2 IMU, Pressure	Online	Recognition	NS	NS	NS	LDA	87.21
Liu & Wang [127]	F S	2 IMU, Pressure, Load cell	Online	Recognition	sliding	15 samples	100Hz	SVM	97.38
Guo et al. [128]	F S R	7 IMU	Offline	Recognition	NS	NS	NS	NNR	2.09° MAE
Qian et al. [129]	F S R	3 IMU, Camera	Online	Prediction	sliding	6 samples	400Hz	CNN	98.5

Data windowing methods

When using machine learning models to classify time-based data, windowing is commonly used to add dimensionality to the data [131]. Data is sectioned to combine sequential datapoints into a single datapoint, so the dimension of time is added to individual datapoints. This can also reduce datasets to reduce processing time. There are two main windowing techniques, non-sliding windows where the next window begins sequentially after the last window, and sliding windows where part of the windows overlap [92].

In the Labarrière et al. [96] review, 30 of the articles used sliding windows, 26 used non-sliding windows, and 2 did not provide information on the windows that were used. Five studies used a section of a stride as the window, the rest used time-based windows. The average time-based window size was 0.247s with the longest being 0.8s and the shortest being 0.05s. The classification accuracies between time-based windows and stride-based windows were similar (mean of the presented accuracies for time-based studies was 95.97% and for stride-based studies was 95.98%). Non-sliding time-based windows performed slightly better than sliding windows (mean of the presented accuracies for non-sliding - 96.92%, sliding – 95.35%).

For the review update, fifteen studies used time-based windows with the shortest being 0.004s [114] and the longest being 8s [109]. Nine studies used windows based on the number of datapoints, with the smallest being 6 datapoints and the largest being 260 datapoints which, considering the sampling rate, translated into time-based windows of 15ms and 1.3s, respectively. The mean window size in time was 1.08s. Eight studies used a stride or the same portion of a stride as their analysis window, three studies analysed individual datapoints and twelve used sliding windows, with the rest not stating the type of windows being used. The accuracies between these three techniques did not greatly vary (mean accuracies for stride was 97.01%, for sliding 97.11%, and for individual datapoints 97.94%).

A range of window techniques have been used to-date and there is no conclusive agreement as to which one gives the best accuracies. Stride-based windows, time-based and individual datapoints all produce similar accuracies. Non-sliding windows produced accuracies slightly better than sliding windows, but those were not significantly higher. Due to the small sample size of most studies and the use of cross-validation to assess accuracies, which could inflate accuracies due to data from the same participant being used to train and test the models, it is inconclusive if windowing technique has an effect on classification accuracies. Therefore, there is a need to explore various windowing techniques, which has not been done in previous literature.

Sensor

In the review update, only one study did not use an IMU [106]. For the remaining studies, the number of IMUs used varied between 8 and 1. Accuracies between studies that used a single IMU and studies that used multiple IMUs or IMUs in conjunction with other sensors did not significantly vary (single IMU: 98.14%, other studies: 97.01%). Pressure or load measuring sensors were the next most-commonly used type of sensor with ten studies incorporating them into their sensing systems. Five studies used cameras or sensors that visually detect the terrain the participant is about to traverse. These studies produced accuracies similar to the mean of all the studies (visual sensors 97.03%, overall mean 97.00%). Three studies used EMG, but always in combination with other sensors.

However, from Labarrière et al. [96], 36 studies used IMUs, of which 7 used IMUs in isolation. Forty-five studies incorporated pressure or load, with 4 using these in isolation. Twenty-one articles incorporated Electromyographs (EMG) and one used Forcemyographs (FMG), but these were always in combination with other sensors. Twenty-one studies used angle encodors but only one used them in isolation. Four studies used lasers or depth cameras to measure distance, but only in combination with other sensors. The mean accuracies of studies that used IMUs were 95.98%, pressure or force 96.36%, angle encoder 96.92%, EMG/FMG 93.04% and laser/depth camera 96.73%. The mean accuracy for using IMUs in isolation was 92.82%, but this was reduced by one study that had an accuracy of 78% using a single IMU which, if excluded, raised the mean accuracy to 95.78%; the only other study that just used a single IMU had an accuracy of 94.1%.

There was not a significant difference in the accuracies produced between the different sensors. A visual sensor, for example a depth camera or laser, would not be practical for real-world use as they need line of sight, thereby restricting what clothes a user could wear, but could be used to train a more practical sensor to reduce data processing. EMG sensors need to be attached to the user's skin which is not practical for long term real-world use, as this would require the prosthetic user to attach the electrodes. In all the studies that used EMG, multiple electrodes were used with the minimum being 4 [117] and maximum being 14 [101]. Notably, asking the prosthetic user to attach electrodes will increase the risk of known issues with EMG, for example artifacts and recording the wrong muscle [132].

Description of Algorithms

A range of different machine learning algorithms were used in all the studies, and they all showed good accuracies for the recorded activities. Prior to discussion of the algorithms used in the papers, descriptions of the most popular algorithms are given below.

Logistic Regression (LR): LR is a statistical model for binary classification. LR uses a sigmoid function to map feature values between 0 and 1, to make predictions of probability between two classes [133]. Due to this LR is regarded as a simple and easy algorithm to implement, but not ideal for multi-class classification [134].

Linear Discriminant Analysis (LDA): LDA is considered an easy algorithm to implement and was used in quite a few studies as a baseline to compare other algorithms against. LDA is a linear classification algorithm. LDA assumes the data is gaussian distributed and that each attribute has the same variance. LDA makes predictions by estimating the probability that an input belongs in a class. LDA can struggle in cases where classes are not linearly separated. LDA makes use of the entire data set to estimate covariance matrices and thus is somewhat prone to outliers [135]. Quadratic Discriminant Analysis (QDA) is an LDA algorithm where the variance for each group varies.

Support Vector Machine (SVM): SVM algorithms were shown to have a slightly better performance than LDA in some studies [96]. The objective of SVM is to find the ideal hyperplane that separates classes with the biggest margins. Separating classes by the biggest margins gives more confidence that a new data point will be classified correctly [136]. SVM does not make assumptions about the data and incorporates a slack variable, thereby allows some overlap between classes [137].

K Nearest Neighbour (KNN): KNN algorithms are similar to SVM and LDA but work by classifying a new data point by seeing which data are similar to that point with a variance defined by the value of K. This means the algorithm is versatile and simple to implement [134] but can become slow with a large number of samples.

Dynamic Bayesian Network (DBN): A DBN is a Bayesian network that models sequences of variables. Bayesian networks are simple graphical models' conditional dependence on edges in a directed graph, and each node corresponds to a unique random variable [138]. DBN has shown better accuracy than LDA [96] and can take transitional possibilities into account to improve classification. Hidden Markov Model (HMM) are similar to DBN, but the entire state of the world is represented by a single hidden state variable [139].

Decision Tree (DT): DT are classifiers that uses a number of if-then-else decision rules to classify data. DT are simple to set up, can handle a range of data types and can be visualised. The main issues relating to DT are that they can be overfit to the data and any change to the input data will create a different tree [140].

Random Forest (RF): RF are made up of several DT models with each tree voting on the class of the input and the class with the most votes is where the data is classified. This reduces the risk of overfitting and provides more flexibility but requires more computational resource [141].

Artificial Neural Networks (ANN): ANN are networks of processing elements that operate on local data and communicate with other elements. The construction of ANN was inspired by the structure of the brain [142]. Many types of ANN have been created but not all are suitable for the type of classification this PhD is looking to conduct because they cannot process time dependent data. ANN are robust and less affected by noise in the data. They are also non-linear and can adapt without user input. But, depending on how many layers and neurons are in the network, they can be computationally expensive. Some ANN employ deep learning and generally require less pre-processing than other machine learning methods [143].

Feedforward Neural Networks (FNN): FNN are ANN that only have connections that pass data forward and do not cycle. FNN are one of the simplest ANN and the first type invented, and they remain one of the most popular ANNs [144].

Convolutional Neural Networks (CNN): CNN are ANN that employ the mathematical operation convolution, which is how the shape of a function modified by another is described by a third function. CNNs use convolution in at least on layer of the network. CNN were designed for image recognition but have been applied to other classification problems [145].

Long Short Term Memory (LSTM): LSTM are a type of recurrent neural network (RNN). RNN are more complex than FNN as they keep information about past inputs to influence future outputs. RNN can have issues with trending towards zero, but LSTM solve this problem by allowing gradients to also flow unchanged if needed [146].

Algorithms used in the reviewed papers.

For the Labarrière et al. [96] review, the three most common used algorithms were LDA (N=15), SVM(N=12) and DBN (N=11). DBN was mainly used by 2 research teams [147][148][149][150][151] [152][153][154][155] where the research teams investigated the same prosthesis or system in all of their articles. Furthermore, across the papers reviewed, three types of neural networks were used, CNN (N=4), ANN (N=5) and LSTM (N=1). KNN, LR, and DT were each used once. Quadratic Discriminant Analysis was used 3 times. Learning From Testing data, Hidden Markov Model, and Entropy Based Algorithm were all used twice. All the algorithms produced good accuracies above 85%.

From the articles identified through the review update, most studies only investigated one machine learning algorithm. Out of the studies that compared algorithms, SVM outperformed LSTM in one study [117] and once produced better accuracies than LDA, QDA and ANN in another [99]. LSTM also produced similar accuracies to CNN in one study [110], but performed better in another, and in the same study produced similar accuracies to a ResNet algorithm, which is similar to CNN but with a wider use of feature extraction [98]. SVM was used in 6 studies, LSTM in 7, CNN in 9 and ANN in 5, with other algorithms only used in a single study.

There is no conclusive answer as to which machine learning algorithm gives the highest accuracies when classifying human movement, as the vast majority of studies only investigated one algorithm. The updated review shows that the use of neural network classifiers has increased in recent years. SVM appears to outperform other classifiers on the same data, this was also found by Jamieson [40], who found that SVM, KNN and LSTM outperformed RF and LDA in classifying terrain being traversed. Chen et al. found that DT classifier outperformed DBN and KNN classifier at measuring joint angles using accelerometers and gyroscopes for rehabilitation exercises, but their findings were based on a small data set [156]. Lum et al. found RF produced better accuracies than SVM and KNN for detecting arm movement for stroke patients [75].

In conclusion, the machine learning algorithm that will produce the best accuracies for a dataset depends on the dataset. The type of data this PhD is concerned with has not been previously classified, so there was no study to help judge which algorithm is best to use. SVM, RF and LSTM performed well in previous studies that have compared different algorithms. LR and KNN are simple algorithms to implement and have shown good accuracies, so could be used to set benchmarks for comparison. LR works best with binary classes so could only be used in walking aid recognition. Importantly, LSTM networks have the capability to keep information which would allow the algorithm to account for a longer time period which will make it more useful than CNN, FNN and ANN for this research. Due to this KNN, LR, SVM, RF and LSTM algorithms will be used for terrain classification. LR, KNN, LR, SVM, RF and LSTM algorithms will be used for walking aid recognition.

Evaluation of classification accuracy

There was no information in the Labarrière et al. [96] review about how accuracies were calculated. Research by Jamieson [157] used both cross-validation and leave-one-out when testing the accuracies of terrain classification, and got different results as to which algorithm works best for the two verification techniques.

Fifteen of the studies identified through the updated review combined all the data that had been collected before splitting it into training and testing groups. Seven of these studies used cross-validation where this split is done multiple times, usually five or ten times, and the mean accuracy is stated. This means that data from the same participant will be in both groups and therefore could compromise the validity of the accuracies in real-world use. Five studies used the leave one out approach, where a participant or group of participants is left out of the training group and instead is solely used to test the algorithm on. This gives a more valid accuracy for real-world use. Out of these studies none used just IMUs, with two combining IMUs with a camera, one combining IMUs with pressure insoles, one combining IMUs with EMG and one using IMUs and measurements for knee and hip angles. Only one of these studies tested on a target population [116], i.e. it tested on a transtibial amputee. This study used an IMU and a camera and produced an accuracy of 93% when classifying between flat terrains, grass and stair use, using a Bayesian neural networks algorithm.

Guo et al. did not look at classifying the terrain but to predict the angle of the terrain the participants were traversing [128]. For this they used mean absolute error of the angle instead of a percentage accuracy.

Most studies have used cross-validation to calculate accuracies, which could be due to the low number of participants in these studies, but leave-one-out has been successfully implemented in research similar to this PhD. To evaluate and compare the accuracies to similar research, both cross-validation and leave-one-out should be used in this PhD.

Filtering

The Labarrière et al. [96] review did not report on filter techniques. From the review update only three studies stated the filtering they used for IMUs data. They all used a Butterworth low-pass filter. Qi et al. [119] used a second-order filter with a cutoff frequency of 5 Hz, Chauhan et al. [109] also used a second order filter but at 6Hz and Li et al. [101] used a sixth order filter at 10Hz. In other studies, Butterworth [158][159][160] and Kalman [73] filters have been used in several studies, with the majority using a low pass filter of around 20Hz [161][162][160] to reduce noise that is not caused by the stride. Most studies normalised their data but the process for this was not routinely mentioned. A second order Butterworth filter has been shown to remove unwanted signal frequencies in similar data to the studies in this PhD. The cutoff frequency can be obtained through frequency analysis of the collected data.

3.3.3 Discussion and conclusions

From these reviews it can be seen that high accuracies for terrain recognition can be achieved using wearable sensors, but real-world use in large target populations has not yet been shown. There is no conclusive agreement on the best algorithm to use for human movement classification. SVM, RF and LSTM have performed well in previous studies, LR and KNN could also be used for benchmark comparisons. There is no conclusion on window type and size. Studies using time-based windows vary in window size from 4ms to 8s, all producing similar accuracies. Some studies have produced similar accuracies splitting the data into strides but the same can be said for studies that have analysed individual datapoints. It has been shown that a single IMU can produce good accuracies, which would meet the requirements set out in Chapter 2 for a system that can be attached to a prosthetic limb and not affect the use of the prosthetic. Most studies did not use a leave one out approach to assess the accuracy of the systems, but doing so would provide a more realistic real-world accuracy.

In Conclusion while various machine learning algorithms—including SVM, RF, and LSTM—have shown high accuracy in activity classification, no single algorithm emerged as better/superior for all types of activity data. Given the unique demands of classifying movement in prosthetic users across diverse terrains, this thesis will compare several models, including SVM, RF, and LSTM, for their ability to handle complex, real-world data. LR will only be used for walking aid recognition. As the data will not be analysed in real time, multiple window techniques will be compared. One where each window is one stride, different time-based windows from 0.1s to 2s and also individual data points. The data will be normalized to reduce bias and a low pass Butterworth filter will be applied to reduce noise.

3.4 Review of Sensors for unsupervised activity classification in the real world

The previous sections detail the review of the literature concerned with algorithms for activity monitoring. In a second step, publications were reviewed with a focus on sensor types used for data collection.By evaluating the strengths and limitations of various sensor types, this review aims to identify the most suitable options for capturing relevant activity data in diverse environments.

Inertial Measurement Unit (IMU)

IMUs consist of accelerometers, gyroscopes and magnetometers combined in a single housing. This means that they do not just give linear acceleration data but also angular velocities and orientation (pitch, roll, yaw). IMUs are one of the most common types of sensors used for activity monitoring. IMUs have become a popular choice due to their non-intrusive nature, portability, low-cost, and ability to capture high-quality inertial data [163].

The Labarrière, et al. [96] review identified three studies that looked only at IMU data. One study mounted an IMU on the prosthesis shank, but they were only investigating cross slope steps. Cross slopes are slopes that go across the body either left to right or right to left, and cross slope steps are used to simulate uneven terrain in a lab environment. They managed to achieve an accuracy of between 59% and 87%, using an LDA classifier [164]. Another paper reported on a study in which two IMUs were attached to a prosthesis to identify ramps and stairs [165]. They did not state the accuracy of the SVM algorithms and had only one participant, so the algorithm will have high bias because there are gait variations across the population. Another study attached IMUs to the shank, ankle and thigh of 11 participants, but only one of these was a prosthesis user [158]. This study had an 82% accuracy at identifying ramps and stairs, but it tested the classifier using a randomly selected 20% of data so there will be bias and no guarantee this accuracy would be achieved with a new participant.

Fourteen studies in the review update used solely one or more IMUs. Five of these used a single IMU, but only one of these incorporated their target population [122]. This study used an IMU attached to the foot to identify flat ground, stair use and ramp walking, to an accuracy of 98.5% although they did not state the verification method used. Three of the other four articles used a shank mounted IMU to classify between flat ground and stair use [110] or flat ground, stair use and ramp walking [112][113]. All these studies achieved an accuracy above 96% using cross-validation.

Ibrahim et al. used the stride length and other spatio-temporal gait parameters to predict the fatigue of MS patients [76]. The prediction was done using a RF regression algorithm and achieved an error of 1.38 on the Borg scale.

Another study mounted IMUs on the chest, hip and ankle of the participants and achieved a good accuracy for activity recognition using a LSTM algorithm but only tested healthy participants [166].

Luo et al. produced a database of IMU data for healthy participants walking over flat ground, banks, grass, cobble, stairs and slopes [159]. Six IMUs were attached to the participants shins, thighs, trunk and one wrist. Hu et al. applied CNN, LSTM and a LSTM structure with an extra global pooling layer to Lou et al. data [167]. The global pooling layer is where each cell learns the global information by getting the distance from other cells at each timestamp. It is therefore able to learn the correlation between different parallel time series. With all the sensors, all classifiers achieved over 90% accuracy and using just one shank IMU LSTM and the global pooling LSTM achieved a similar accuracy. Data for left and right shank were not combined, and this might have improved accuracy. Dixon et al. applied a FNN and attained a 97% accuracy in a 7-fold cross-validation test [168].

A few studies used the components of an IMU separately. Redfield et al. found that with one accelerometer attached to the ankle of the prosthesis 90.1% accuracy could be achieved with a DT classifier at identifying sitting, standing, moving, and doffing [67]. Lum et al. found that RF and radial basis function SVM classifiers produced the best accuracies at measuring the functional use of an arm from wrist worn accelerometer data [75]. Chen et al. found that DT classifier outperformed DBN and KNN classifiers at measuring joint angles using accelerometers and gyroscopes for rehabilitation exercises, but their findings were based on a small data set [156].

Accelerometers

Accelerometers have shown to be able to classify distinct activities like sitting, standing, walking, and lying [169], but are not commonly used for terrain recognition. As mentioned in Chapter 1, Jamieson [40] used a thigh mounted accelerometer to try and recognise activities of lower limb amputees in free living conditions. Comparing SVM, KNN, RF, AB, NB, LSTM and LDA, it was found that SVM and LSTM performed the best. LSTM produced accuracies of 78.43% when assessing using a 5-fold cross-validation on all collected data and SVM produced accuracies of 56.68% when using a leave one out approach. These accuracies are not as high as ones achieved with different sensors, but the study did incorporate 4 lower limb prosthetic users as well as eight healthy participants.

Pressure sensors

Pressure sensors were used in three papers from the review. One mounted the pressure sensor inside the socket of the prosthesis and achieved a 97% accuracy at identifying brisk walking and stairs with a multi-layer FNN, but they only used one participant so will have high bias [161]. The other two studies used pressure-sensing insoles. One achieved a 99% accuracy with recognising sitting, standing, walking, stairs and stepping over obstacles with a DT classifier but had only one test participant [170]. The other looked at the same locomotion and attained a 98% accuracy, with K-fold cross validation, using LDA [171].

Studies combining different sensors

A few studies have combined sensors to recognise locomotion. One study attached an accelerometer and gyroscope to a transfemoral socket and a pressure sensing insole to identify walking on stairs and ramps [172]. Using all sensor data and a HMM, a 95.8% accuracy was achieved. This study combined amputee and non-amputee data to achieve this accuracy, and found that the data of amputees and able-bodied participants produced similar patterns but with different amplitudes. Another study attached IMUs to the back of the shank and shoe combined with a pressure sensing insole and produced a similar 3.6% error using LDA and QDA classifiers [74]. One further study implanted a load cell and IMU to the shank of a prosthesis; they found that SMV gave the best accuracies but was slower than LDA and QDA classifiers [173]. The final study combined EMG, IMU and load cell data for real time locomotion recognition and achieved a 95% accuracy [162].

Nineteen studies from the review update combined sensors. Only one of these studies did not use an IMU, and combined EMG with pressure data [106]. Nine of the eighteen other studies used pressure or load incorporation with IMUs, the average classification accuracy of these studies was 96%. Only one of these studies though recruited any participants from the target population [107]. This study used pressure insoles with a thigh and foot mounted IMU to classify between flat ground, stairs use and ramp walking. Five studies use a visual sensor with an IMU, but a visual sensor would not be practical for real-world clinical use. Three studies incorporated IMUs with EMG and one combined IMUs with joint angle measurements.

Camargo et al. showed that IMUs and goniometers can be used to calculate walking speed, stair height and ramp incline to a high degree of accuracy [73]. The system used four IMUs, on the foot, shank, thigh, and trunk, and 3 goniometers, on the ankle, knee, and hip. The shank IMU reduced the error the most when calculating walking speed, while the goniometers reduced the most errors in calculating the step height and ramp angle. Without the goniometers, the error of the step height was about 5cm and the error of the ramp incline was about 3.5°, but with the goniometer the errors improved to 1.29cm and 1.25 degrees. The shank IMU could detect a ramp but could not accurately measure the incline. The classifiers proving most successful in terms of reducing errors were feedforward neural network and DBN with a Kalman filter. Notably, goniometers measure angles so will not be effective on a prosthesis with a fixed ankle joint.

Make of sensors and specification

A range of different makes of sensors have been used and shown to be valid for activity monitoring, all with similar specifications. This shows that the make of the sensors used for this research is not as important as the specification of the sensor. There was a range of data collection frequencies of the sensors used, from 50Hz [70] to 500Hz [150], but the majority of sensors used 100Hz [158][74][173][162][159][167][168][76][67][166]. Although walking has been shown to be accurately measured at lower sampling frequencies, and the ActivPal uses a default 20Hz, recording data at a higher frequency will allow tests to be carried out to see which frequency will give the most accurate predictions.

Discussion and conclusions

This review of sensors showed that IMUs have been the most frequently used sensors across all studies [165][158][74][173][162][159][167][168][73][76][166] and shown to be able to classify different types of locomotion with good accuracy. This PhD aims to create a system that only requires attachment of sensors to the prosthesis so that the participants do not need to attach anything to their body. In the subsequent work, data will be collected to enable exploration of accuracy versus data redundancy, hence initially sensors will also be attached to the participants' body. Copying the IMU locations used in the study by Luo et al. [159] would give adequate data and will give data for prothesis users that can be compared to able body participants. Pressure sensors and load cells measure similar data and have regularly been used in combination with IMUs [74][173][162]. Individually they have been shown to accurately identify stairs and obstacle avoidance [161][170] and provide an easy method for recording wear-time. As discovered in Chapter 2, the simpler the system the more likely it will be used by prosthetic users. Considering this, although measuring load has been shown to produce good accuracies, measuring this data will add to the complexity of the system for the prosthetic user. Just using IMUs has shown to produce equally good accuracies with terrain classification as studies that have incorporated load data, and IMUs are able to measure the kinematic data that should be able to classify walking aid use. Notably, the sensors will have to have a data collection frequency of at least 100Hz.

3.5 Summary of findings

Through review of the literature, this chapter identified several gaps and provided useful insights :

- Only 2 studies specifically focus on recognising walking aid use with wearable sensors. The available studies use wrist-worn accelerometers, which may not be practical or acceptable for long-term use by prosthetic users.
- The majority of studies in activity classification and prosthetic use have been conducted with healthy participants rather than prosthetic users. Studies that have prosthetic users often have very small sample sizes.
- Very few studies consider uneven or complex terrains like gravel and cobblestones, which prosthetic users encounter in real-world environments.
- A variety of data windowing techniques have been used in the past but there is no definitive agreement on the optimal window type.
- There is no conclusive agreement on the best machine learning algorithm for terrain or walking aid use classification as such is dependent on the dataset collected.
- IMUs have previously been shown to be able to classify terrain and should be able to measure the kinematic data that could be used to classify walking aid use.

This chapter established that whilst a vast amount of algorithms and sensor sets have been explored in connection with activity classification, there remains a need for further research to develop a real-world monitoring system capable of accurately recognising a wide range of activities in prosthetic users, including walking on uneven or unstable terrains with and without a walking aid. To develop such a system, research will need to compare machine learning classification algorithms and aspects of these algorithms, for example data windowing techniques, to investigate the techniques that produce the highest accuracies.

Informed by this literature review, the next chapter (Chapter 4) is concerned with the design of a system with the following specifications:

- **IMU will be used to capture activity data.**
- **IMU will capture data at a sampling frequency of 100Hz.**
- **KNN, LR, SVM, RF and LSTM algorithms will be compared for terrain classification.**
- **KNN, LR, SVM, RF, LSTM and LR algorithms will be compared for walking aid recognition.**
- **Time-based window (0.1s to 2s), stride-based windows and analysing individual datapoints will be compared.**
- **A low-pass Butterworth filter will be used to filter the raw data.**
- **Data will be normalized to reduce bias.**

Chapter 4

System Design: Classification of terrain and walking aid use using real-world data.

4.1 Background

In the previous chapter (Chapter 3), review of the literature concerned with sensors and algorithms informed a set of system specifications. The **purpose** of this chapter was to design the system informed by these specifications. The overall aim is to create a system capable of accurately classifying user activities, terrains, and walking aid use in diverse environments. This initial design will be tested across a range of terrains and with different participants to assess baseline performance.

4.1.1 Aims of Chapter 4

- Using IMUs, collect data on 20 lower limb prosthesis users traversing different terrains (stairs, slopes, grass, gravel, and cobblestones) with and without walking aids.
- Investigate the use of machine learning algorithms to create models to classify terrain traversed and walking aid use for lower limb prosthetic users.

4.2 Methodology

The Methods are described with regard to recruitment and participants, use of instrumentation, details of the protocol, data collection, and analysis methods. Furthermore, methods for data pre-processing are discussed with respective underlying rationales for processing choices made.

The data collection will involve lower limb prosthetic users conducting trials where they will traverse a set terrain with or without a walking aid with sensors attached to their body and prosthesis. The collected data will be processed and used to train and test classification models with the aim to produce a model that can identify the terrain and if the participant is using a walking aid from the sensor data.

Ethical approval was granted to collect data on campus, and also for collecting data at a location convenient for the participants (Ethical approval numbers for University and IRAS Ethics: 4743 and 314743, Appendix D.2 and D.3).

4.2.1 Terrain

Incidents where amputees negotiate stairs, ramps and uneven terrain were mentioned during the interviews in Chapter 2 as being of interest to clinicians. Uneven terrain was specified as terrain that would be likely to produce a different foot-to-shank angle on each step. Grass and unstable terrain like gravel or sand were deemed common terrains that someone could traverse depending on where they lived, and were mentioned by participant 4 in Chapter 2, so those were also included in the data collection.

At the University of Salford, stairs, steps, slopes, grass, gravel, cobble stones and uneven pavements were available to be used for the data collection. As shown in Figure 4.1 there was a clear path that lead from one terrain to the other in the order of gravel, stairs, steps, slopes, grass, uneven pavement and cobblestones. Due to this arrangement, this was the order in which the different terrains were traversed by the participants to enable a single recording without breaks, but the order was reversed for half the participants to reduce effects of fatigue on the later-traversed terrains. Uneven terrain and cobble stones were considered together as uneven terrain. Data on both stairs and steps were collected due to differences in heights and widths which require a different gait to traverse one or the other, consideration of which aids in building robustness into the algorithm. All these terrains met the requirements set out by the clinical experts in Chapter 2.

Fig. 4.1 Map of pathway used for real-world data collection at the University of Salford.

Chapter 5 uses a motion capture system to create virtual IMUs to see if sensor placement on the limb would affect the accuracy of the classification algorithms. For this project lower limb prosthetic users traversed stairs, a ramp, artificial cobble stones and flat terrain. Details about these terrains can be found in Chapter 5. IMU data was collected during that data collection that was also used in this classification model development.

To increase participant numbers, data collection was expanded to locations convenient for the participants but which had some of the required terrains. This was generally near the participants home or work. This not only helped with recruitment but also further increased the robustness of the algorithms due to the use of additional environments and different terrains. Table 4.1 displays the location of the data collection for each participant.

The issue that arose from this type of data collection was, however, that in some locations not all the required types of terrain were available. This meant that for some participants data for some types of terrain were not collected.

Types of terrain were also excluded from the data collection for individual participants if they did not feel safe traversing them. This was sometimes weather dependent, for example one participant did not wish to walk on grass while it was still damp after rain. Furthermore, for one of the participants there were technical issues with the sensors that affected the data collection session, so that not all the different terrains were able to be recorded. Table 4.2 presents the terrains data was collected on for each participant.

4.2.2 Participants

To ensure the participant group was the most relevant to the issues raised in Chapter 2, i.e. distinguishing between a K2 and K3 patients, the inclusion and exclusion criteria for participant recruitment were designed to exclude K0, K1, and K4 patients. Because, not all prosthetic users know their K level the exclusion criteria also includes criteria that would exclude them from being a K2 or K3, for example participating in active sport or regular use of a wheelchair. The inclusion criteria were also written to ensure participants would be comfortable using a walking aid and traversing most if not all the required types of terrain. The inclusion and exclusion criteria were:

Inclusion Criteria:

1. Be a transtibial or transfemoral amputee
2. Be able to walk on a lower limb prosthesis (i.e. mobile with prosthesis)
3. Have experience using a walking aid
4. Be comfortable with all or some of the following: climbing/ descending stairs/ramps and walking on uneven ground
5. Be able to provide informed consent

Exclusion Criteria:

1. Have been classified as a K0, K1, or K4 patient (K0 and K1 are prosthetic users that are unable to walk on their prosthetic outside the home, and K4 are prosthetic users that participate in active sport on their prosthesis)
2. Use primarily a wheelchair rather than walking on their prosthesis

3. Regularly participate in an active sport on their prosthesis (football, running, etc.)
4. Not able to understand written and spoken English
5. Having a positive test result (LFT or PCR) for Covid or having symptoms of Covid without a negative PCR test
6. The study finishes while they decided to take part

4.2.3 Recruitment

To recruit for this study, a few methods were used to ensure there is a diverse range of participants including advertisements on social media, advertising through prosthetic user-focused organisations, university professional patients, and NHS prosthetic clinics. The social media advertising was conducted on X. The advert (appendix F.1) was shared by the clinical and academic prosthetic organisations and professionals in the United Kingdom. The Limbless Association and Manfit shared the advert with their members. Professional patients at the University of Salford and the University of Strathclyde were asked if they would like to participate. Manchester University NHS Foundation Trust and Portsmouth Hospital University NHS Trust put up posters to advertise the study (Appendix F.1) and shared the study with patients that met the inclusion/exclusion criteria when they visited the clinic.

Twenty participants were recruited, provided informed consent, and participated in the study. Eleven were transtibial (TT) and 8 were transfemoral (TF), with one participant being a bilateral amputee with a TT and TF prosthetic. For this study no participant was excluded due to demographics as having more data can help to make machine learning algorithms more robust. Table 4.1 displays the information for each participant. For Participant 7, who was a bilateral prosthetic user, the IMU set up was changed to IMUs on both prosthetic shanks and both thighs. This was done under consideration that the shank and thigh provide more important information than the trunk, and data could be collected for both a TT and TF leg at the same time without the participant needing to conduct twice the number of trials which could have caused the participant fatigue. Participant 7 was considered as two participants participant 7 for the right TT prosthetic and 21 for the left TF prosthetic. The data for these two participants will be from the same person but will focus on different strides from different prosthetic types so will not be similar and will not affect the classification accuracies.
Not all the participants were assessed on every terrain, some participants also did not want to use the walking aid, and one participant only felt comfortable conducting the

Table 4.1 Descriptive patient data.

Participant no.	Age	sex	Transtibial (TT)/ Transfemoral (TF)	Location for data collection
P1	85	Male	TT	University of Salford
P2	63	Male	TT	University of Salford
P3	69	Male	TF	University of Salford
P4	80	Male	TF	University of Salford
P5	74	Male	TT	University of Salford
P6	69	Male	TT	University of Salford
P7/P21	66	Male	TT/TF	University of Salford
P8	56	Male	TT	University of Salford
P9	53	Male	TF	University of Salford
P10	59	Male	TF	University of Salford
P11	72	Male	TT	University of Salford
P12	64	Male	TF	University of Salford
P13	69	Male	TT	Participants home
P14	44	Male	TF	Participants home
P15	33	Female	TT	Participants home
P16	67	Male	TF	Participants home
P17	62	Female	TT	Participants home
P18	61	Male	TT	Participants work
P19	80	Male	TF	Participants home
P20	65	Male	TT	Participants home

study with a walking aid used at all times. Data collection constraints for each participant are displayed in Table 4.2.

The only type of personal walking aid used in the data collection was a walking stick as it was the personal walking aid all the participants used in their everyday lives. Handrails for stairs and slopes were also classified as a walking aid as as it was deemed that a participant would use it to offload and therefore could change their gait.

Table 4.2 Test conditions for each patient. "wi": with walking aid, "wo": without walking aid.

Participant no.	Flat	Grass	Up Stairs	Down Stairs	Up Slope	Down Slope	Unstable	Uneven
P1	wi/wo	wo	wi/wo	wi/wo	wi/wo	wi/wo	wo	wi/wo
P2	wi/wo	wo	wi/wo	wi/wo	wi/wo	wi/wo	wo	wi/wo
P3	wi/wo		wi/wo	wi/wo	wi/wo	wi/wo	wi/wo	wi/wo
P4	wo	wo	wo	wo	wo	wo	wo	wo
P5	wi/wo				wi/wo	wi/wo		wi/wo
P6	wi/wo	wi/wo	wi/wo	wi/wo	wi/wo	wi/wo	wi/wo	wi/wo
P7	wi/wo		wi/wo	wi/wo	wi/wo	wi/wo	wi/wo	wi/wo
P8	wi/wo		wi/wo	wi/wo	wi/wo	wi/wo		wi/wo
P9	wi/wo	wi/wo	wi/wo	wi/wo	wi/wo	wi/wo	wi/wo	wi/wo
P10	wi/wo	wi/wo	wi/wo	wi/wo	wi/wo	wi/wo	wi/wo	wi/wo
P11	wi/wo	wi/wo	wi/wo	wi/wo	wi/wo	wi/wo	wi/wo	wi/wo
P12	wi/wo	wi/wo			wi/wo	wi/wo	wi/wo	wi/wo
P13	wi/wo	wi/wo	wi/wo	wi/wo	wi/wo	wi/wo	wi/wo	
P14		wo	wo	wo	wo	wo		wo
P15		wi/wo	wi/wo	wi/wo	wi/wo	wi/wo		
P16	wi	wi	wi	wi	wi	wi		
P17		wo	wo	wo	wo	wo		
P18		wo	wo	wo	wo	wo		
P19		wi/wo	wo	wo			wi/wo	

4.2.4 Sensor

In Chapter 3, IMUs were identified to be the most suitable sensors for real-world data collection. Luo et al. [159] used 5 IMUs on healthy participants, one on each shin and thigh and one on the trunk, and Hu et al. [167] produced good accuracies for terrain recognition only using one shank IMU from the healthy dataset collected by Luo et al. Encouraged by Luo's results, and because the prosthetic leg is of most interest and the ideal system would only need to be attached to the prosthesis and not the patient's body, the IMU placement in this study focused on the prosthetic leg. A total of 4 IMUs were used for the data collection, placed on the prosthetic shank, the thigh of the prosthetic leg, the (other)anatomic shank, and the trunk, Figure 4.2.

The placement sites—on the prosthetic shank, the prosthetic thigh, the anatomic shank, and the trunk—were chosen to capture comprehensive motion data relevant to distinguishing gait patterns for terrain classification and walking aid recognition, while minimizing the need for body-mounted sensors.

Fig. 4.2 IMU placements

The specific commercially-available IMU chosen for the data collection were Xsens Awinda IMUs since they met the requirements set out in Chapter 3, of being able to record data at 100Hz and have been prove to be able to capture human walking data as they were the IMUs used by Luo et al. [159].

4.2.5 Study procedure

The procedure for the study was the same wherever the trials were conducted. After establishing that the participant was comfortable conducting the trial, the procedure was as follows:

- Participant confirms they are ready.
- Start video recording.
- Start IMU recording.
- Participant given the signal to start.
- Participant stamps their prosthetic leg on the floor twice, providing a clear start signature on the signal.

- The participant conducts the trial.
- The participant finishes the trial and stands still.
- IMU recording is stopped.
- Video recording is stopped.

The participants were asked to stamp their prosthetic leg twice to make it easier to align the ground-truth video with the IMU data.

The data collection took place between November 2022 and March 2024.

4.2.6 Data Filtering

A low pass Butterworth filter was used to filter the data. To calculate the correct cut off frequency, a frequency analysis was carried out. This involved calculating the Fourier transform of the data and seeing where the main frequencies, i.e. where the main information content, occurs. Each participant conducted multiple trials on each terrain with and without and with out a walking aid, creating a large overall database. Due to the size of the dataset the analysis was not run on all trials but on 20 which were quasi-randomly selected, ensuring that each type of trial was included for one TT and TF participant. The trials selected are displayed in Table 4.3. Due to the length of the data collection period and to ensure the data processing and analysis would be complete in the timeframe of the research, the frequency analysis was carried out before all the participants had conducted the study. Participant 15 was therefore the last participant included in the frequency analysis. The analysis was carried out on the raw sensor data and the results for the different measures were all similar with the main frequencies being below 4Hz. Figure 4.3 presents the resultant acceleration analysis. To ensure all important frequencies were included, a 5Hz low pass Butterworth filter was used, as stated in Chapter 3. Similar to Luo et al. [159] a second order filter was chosen.

Table 4.3 Trials selected for frequency analysis. Participant number and trial number displayed. WA = walking aid

Gravel without WA	Gravel with WA	Stairs without WA	Stairs with WA	Slope without WA	Slope with WA	Grass without WA	Grass with WA	uneven without WA	uneven with WA
P1 1	P11 1	P6 2	P15 3	P1 2	P13 1	P8 2	P6 1	P8 1	P6 1
P3 1	P9 1	P12 2	P10 1	P4 1	P10 1	P4 1	P10 1	P14 1	P9 2

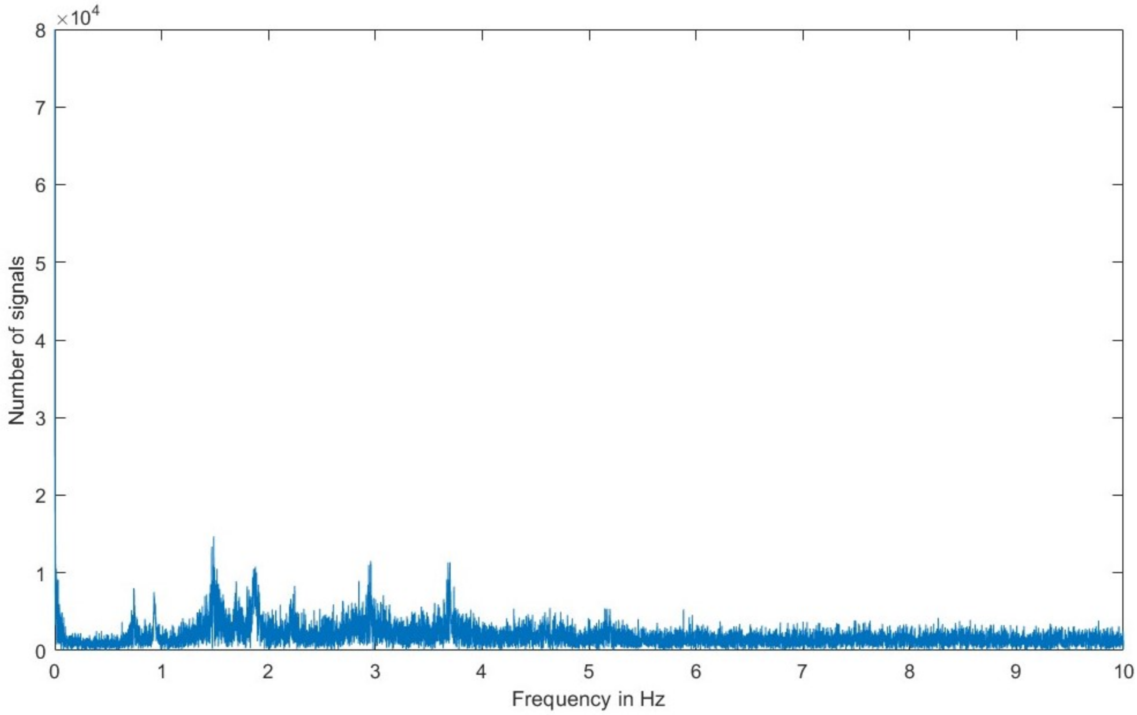

Fig. 4.3 Frequency analysis of resultant acceleration signals. Spikes that identify the main frequencies can be seen at around 1.5Hz, 1.9Hz, 3Hz and 3.75Hz.

4.2.7 Stride identification

The data were analysed using the different window methods identified in Chapter 3. One of these methods is to use a whole stride as an analysis window. To achieve this, the data had to be split into different strides. Also, to label the data, individual strides had to be identified.

No method that has previously been used has looked at the precision of a stride count method to count a stride at the same point in a stride every time. The accuracies previously used to test a stride count method have been the number of strides. As it could be crucial for the accuracy of the classification to analyse the same portion of a stride, analysis was carried out to determine the best stride count method, looking at the accuracy of the number of strides and the precision of where the stride was counted on each stride. Due to the size of the dataset, not all trials were included in this analysis; the trials included were the same as for the frequency analysis as this included one trial for every trial type. For each trial the video was used to count the number of strides, and this was taken as the ground truth. The recorded acceleration and gyroscope data for each of the 3 axes, as well the the resultant magnitude and sum of the three components of acceleration for the prosthetic shank (PS), thigh (TH) and trunk (TR) IMUs were plotted. A 'find-peaks' algorithm was used to count the strides, with the minimum distance between peaks set at 0.8s and the minimum peak height allowed to be adjusted for each measure to achieve the best result. The find-peaks algorithm finds the point where the highest signal occurs within a set portion of time. Table 4.4 presents the percentage errors for all the trials for all the measures, and Figure 4.4 visualises the IMU axis. The percentage error for stride count was calculated as the sum of the difference in stride count between the identified strides and the actual stride count for each trial divided by the total number of actual strides.

$$z = \frac{\sum |S - I|}{\sum S}$$

z = percentage error of stride count, S = actual number of strides per trial, I = number of strides identified per trial

The percentage error for the stride precision was calculated as the sum of the difference between number of strides identified in the same location (the same peak in measurement during each stride shown in Figure 4.5) and the number of strides identified for each trial divided by the total number of strides identified.

$$x = \frac{\sum |I - N|}{\sum I}$$

x= error in stride precision, I = number of strides identified per trial, N = number of strides identified in the same location of the stride per trial

Table 4.4 Stride count % errors. TH: Thigh, PS: Prosthetic shank, TR: Trunk. The PS gyroscope in the Y axis (red) was the measure chosen to identify strides.

Measure	% error of stride count	% error of stride precision
TH resultant acceleration	2.36	6.83
PS resultant acceleration	11.54	10.93
TR resultant acceleration	7.20	6.21
TH sum of accelerations	10.67	11.91
PS sum of accelerations	46.40	6.74
TR sum of accelerations	17.99	23.30
TH acceleration X axis	6.82	21.55
TH acceleration Y axis	11.66	20.89
TH acceleration Z axis	4.09	9.54
TH gyroscope X axis	6.82	26.99
TH gyroscope Y axis	2.23	2.75
TH gyroscope Z axis	4.09	11.87
PS acceleration X axis	4.22	12.34
PS acceleration Y axis	5.21	17.42
PS acceleration Z axis	6.95	21.95
PS gyroscope X axis	7.94	15.58
PS gyroscope Y axis	3.85	2.03
PS gyroscope Z axis	3.60	9.59
TR acceleration X axis	4.47	19.32
TR acceleration Y axis	4.47	21.21
TR acceleration Z axis	5.46	17.90
TR gyroscope X axis	6.45	15.49
TR gyroscope Y axis	3.60	1.91
TR gyroscope Z axis	5.96	17.45

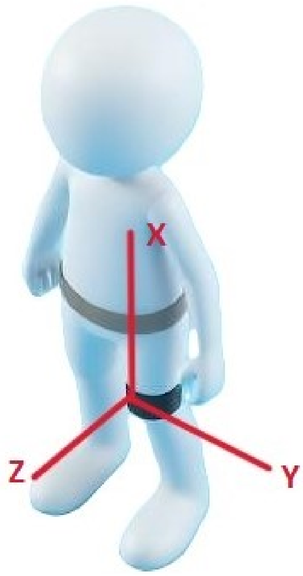

Fig. 4.4 Visualisation of IMU axis

The sum of prosthetic accelerations was found to produce a 46.4% error for stride count which was the worst of all the measures. Trunk gyroscope 'TR' in the Y axis produce the best precision (1.91% error) and a good accuracy (3.6% error), and the thigh gyroscope in the Y axis 'TH' produced the best stride count accuracy (2.23% error) and a good precision (2.75% error). As identified in Chapter 2, an ideal system would be prosthetic-mounted only and not need for a patient to connect a device to their body, so the prosthetic shank gyroscope in the Y axis was chosen as it produced a good precision at 2.03% and a good accuracy at 3.85% which was only slightly worse than the trunk gyroscope in the Y axis. Figure 4.5 plots the resultant acceleration, gyroscope data for the Y axis for the prosthetic shank IMU and force plate data for Participant 5 on flat terrain. It is clear from the plot that there is one clear peak per stride for the Y axis gyroscope data, whereas for the resultant accelerations there are multiple peaks per stride. This is why the resultant acceleration has a higher precision error than the Y axis gyroscope data. The force plate data shows when heel strike occurs, from that it can be seen that the peak that identifies a stride for the gyroscope data in the Y axis is in the middle of the swing phase of the stride.

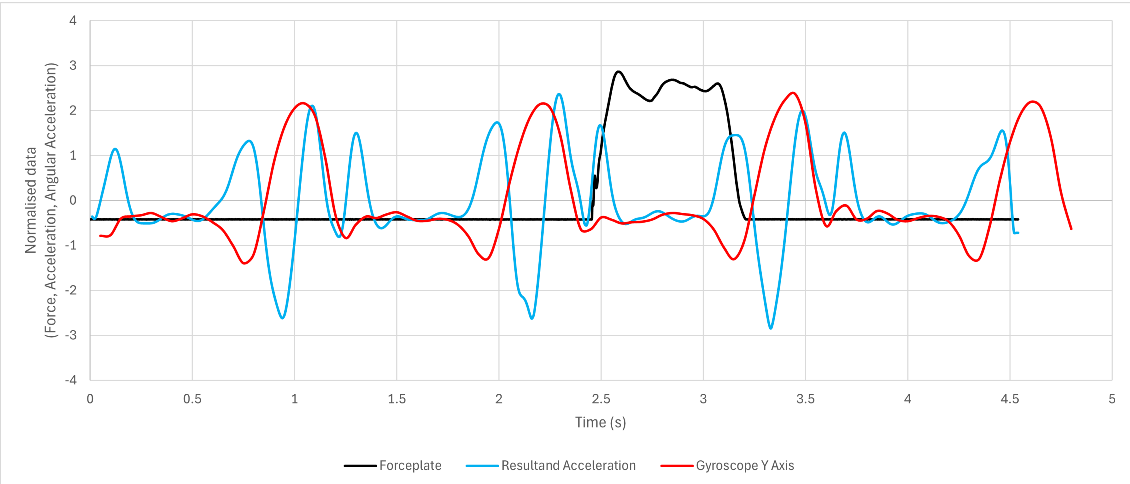

Fig. 4.5 Comparison of strides as defined by different variables. Black line: Force plate data representing the ground reaction force during each stride. Blue line: Resultant acceleration from the prosthetic shank IMU, indicating the total acceleration experienced across all axes during the stride. Red line: Gyroscope Y-axis data from the prosthetic shank IMU, representing the rate of angular change in the medio-lateral direction during each stride.

With the data labelled into strides, the 'stomp' strides at the beginning of the trial and the last partial stride were removed to ensure only complete walking strides were used for analysis.

4.2.8 Labelling

As a stride was counted in the swing phase of a stride, the terrain the foot would land on during the stride was labelled for the whole stride. Transition strides were not uniquely identified. To create training data, the terrain labelling was done manually. The video data was matched to the IMU data, using the stomp and stride count information, and then the terrain. It is acknowledged that this could be a source of error as it required human input for every stride, and a miscount could mislabel the terrain, but it was the only option available for this study. All trials were checked manually by the author after labelling to try and reduce errors to a minimum.

4.2.9 Normalisation

The data were normalised, but as previous studies do not comment on how they have normalised their data, two normalisation methods were tested in this research. The first was normalising the data to compare to the person's mean for each measure, and the other was to normalise each stride individually against the mean for that stride. Raw data that had not been normalised were also tested.

4.2.10 Analysis windows

From Chapter 3, different window sizes and types have previously been used in similar research, but there is not a conclusive agreement on the best method. For this study, six time-based windows were compared as well as using a whole stride as a window and classifying individual datapoints. When the data are processed in the window that consist of a whole stride, the number of datapoints per stride have to be the same, so the strides were resampled to ensure their same size.

4.2.11 Sampling rate

The IMUs record at a sampling rate of 100Hz. Previous studies have used a sampling rate as low as 10Hz to accurately classify terrain [116]. There is no consensus on the best sampling rate to use in this kind of analysis. Nyquist theorem says that the sampling rate has to be twice the largest frequency, to ensure aliasing does not occur [174], so because the data were filtered at 5Hz, the sampling rate cannot be lower than 10Hz. Lower sampling rates require less computational power to process, so in this research four sampling rates were tested to see if the classification accuracy is affected as the sampling rate is reduced. The sampling rates simulated were 10Hz, 20Hz,50Hz and 100Hz. When the data had been split into windows that consisted of a whole stride, the data were resampled into four different numbers of datapoints per stride-window. The number of datapoints per window were 10, 20, 50 and 100. These reflected the sampling rates. As shown in Figure 4.6, the majority of strides were at a cadence of less than 1 stride per second, so resampling to 10 samples per stride will not reduce the sampling rate below 10Hz.

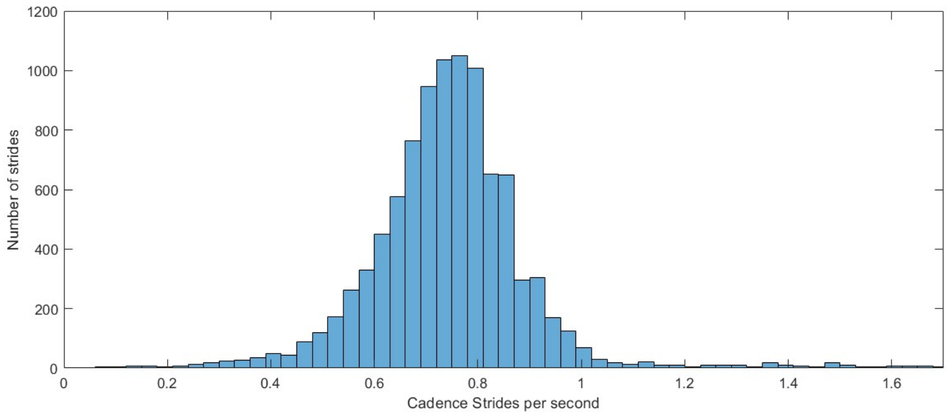

Fig. 4.6 Histogram for strides at different cadences, for all data recorded. The majority of strides are at a cadence lower than 0.8 strides per second.

4.2.12 Features

In machine learning classification features are the variables used to create the classification model. The features used for both terrain classification and walking aid recognition are listed in Table 4.5. The cadence was calculated from the stride time for each individual stride. The percentage along the stride was also calculated for each individual datapoint.

The delta velocity is the change in velocity between datapoints, acceleration is the delta velocity divided by the delta time, if the delta time is constant the delta velocity should be the same but in proportion with the sampling rate.

Table 4.5 Features for terrain and walking aid use classification.

Feature	Units	Description
Acceleration X	m/s^2	Acceleration in the vertical direction (w/gravity)
Acceleration Y	m/s^2	Acceleration in the medio-lateral direction (w/gravity)
Acceleration Z	m/s^2	Acceleration in the anterior-posterior direction (w/gravity)
Free Acceleration X	m/s^2	Acceleration in the vertical direction (w/o gravity)
Free Acceleration Y	m/s^2	Acceleration in the medio-lateral direction (w/o gravity)
Free Acceleration Z	m/s^2	Acceleration in the anterior-posterior direction (w/o gravity)
Gyroscope X	rad/s	Rate of turn along the vertical direction
Gyroscope Y	rad/s	Rate of turn along the medio-lateral direction
Gyroscope Z	rad/s	Rate of turn along the anteriorposterior direction
Magnetometer X	a.u.	3D magnetic field in the vertical direction
Magnetometer Y	a.u.	3D magnetic field in the medio-lateral direction
Magnetometer Z	a.u.	3D magnetic field in the anteriorposterior direction
Velocity X	m/s	Delta_velocity (dv) in the vertical direction
Velocity Y	m/s	Delta_velocity (dv) in the mediolateral direction
Velocity Z	m/s	Delta_velocity (dv) in the anteriorposterior direction
Resultant Acceleration	m/s^2	Resultant accelerations (w/gravity)
Resultant Free Acceleration	m/s^2	Resultant accelerations (w/o gravity)
Cadence	Strides/s	Number of strides per second
Percentage of Stride	%	Percentage of the stride where that datapoint occurs

4.2.13 Analysis

There are two methods of testing the accuracy of classification algorithms. The first being cross-validation which is where the whole data is split into multiple groups, and each is tested while the other groups are used to train the algorithm. For this research, 5-fold cross-validation was used. The other method is leave-one-out, where a whole participant is left out to be tested and the rest of the participants are used to train the algorithm. Due to the number of methods being tested and to reduce computational time, instead of testing individual participants, participants were split into 5 groups, and for each trial four of the groups were used to train the model and one was used to test. This method is called leave-some-out. The participants were quasi-randomly assigned to groups using a random number generator, but it was ensured that at least one TF and two TT participants were in each group. With 21 separate sets of participant data for shank and thigh trials, four groups had four participants and one had five. For the trunk trials where Participant 7 and therefore 21 did not have a sensor, there were four groups of four and one of three. Table 4.6 displays the numbers of the participants in each group.

Table 4.6 Participant numbers and numbers of participants in each group

Group	TT participant numbers	TF participant numbers	Number of participants in each group (TT,TF)
1	P6, P11	P3, P16	4 (2,2)
2	P7, P8, P17	P14	4 (3,1)
3	P1, P5, P15	P19, P21	5 (3,2)
4	P2, P13	P4, P12	4 (2,2)
5	P18, P20	P9, P10	4 (2,2)

4.2.14 Algorithms

As concluded in Chapter 3, four machine learning algorithms were tested for the terrain classification. The four algorithms were SVM, KNN, RF and LSTM. For the walking aid use recognition LR was also tested because LR models look at binary results such as walking aid use/non-use and give a baseline accuracy that other models can be compared against.

Optimisers were run for each algorithm to tune the hyperparameters for each model. The optimisers work by creating classification models with altering the hyperparameters for each model and calculating the loss function. The optimisers try and find the hyperparameters that produce the lowest loss function. All the optimisers were run on 100Hz data and 0.2s windows as well as data resampled to windows of a whole stride with 100 datapoints. The optimisers were run on all the data collected for all participants. There was little to no difference between the optimised outcomes for both window methods for all algorithms. Where there was a difference, a mean value for the hyperparameter was used.

Support Vector Machine (SVM)

The optimiser adjusts different hyperparameters to create a model that minimises the loss function. The hyperparameters that were adjusted were; box constraint, kernel scale and coding method. The hyperparameters found to be the best for terrain classification SVM trials were; box constraint – 969.43, kernel scale 56.59 and one vs all coding. For walking aid recognition SVM trials the best hyperparameters were; box constraint – 574.01, kernel scale 45.609 and one vs one coding.

K-Nearest Neighbor (KNN)

As with the SVM algorithm, an optimiser was run on the same window methods. The hyperparameters that were optimised for KNN were number of neighbours and how the distance is calculated. The best hyperparameters for terrain classification KNN trials were; number of neighbours – 1, how distance was calculated – Euclidean distance. For walking aid recognition KNN trials the best hyperparameters were; number of neighbours – 1, how distance was calculated – Mahalanobis distance.

Random Forest (RF)

The hyperparameters that were optimised for the RF models were number of trees, number of predictors and minimum leaf size. The best hyperparameters for the terrain classification were: number of trees – 100, number of predictors – 7, minimum leaf size - 1. For the walking aid recognition the best hyperparameters were: number of trees – 350, number of predictors – 45, minimum leaf size - 7.

Long Short Term Memory (LSTM)

LSTM are types of recurrent neural networks with LSTM layers. Most LSTM have only one LSTM layer, but multiple LSTM layers can produce deeper learning which could produce a more accurate classifier [175]. To see if more LSTM layers improve the accuracy of the terrain and walking aid use classifier, trials were run with one, two and three LSTM layers, the results of which are displayed in Table 4.7 for terrain classification and Table 4.8

for walking aid recognition and visualised in Figure 4.7 for terrain classification and Figure 4.8 for walking aid recognition. The analysis was run on data split into stride windows of 10 datapoints per stride and used 5-fold cross-validation to assess the accuracies. There was no significant difference between using one (M=72.39%, SD=1.73%) and two (M=72.49%, SD=1.19%) LSTM layers for terrain classification using a Dunn's test (0.1% difference in mean accuracy, p value = 0.96), whereas there was a significant decrease in accuracy when using three layers (M=66.38%, SD=1.56%) (6.01% difference in mean accuracy, p value = 0.011) as compared to one layer from a Dunn's test. One layer (M=6.4s, SD=0.49s) was also significantly quicker to run than two layers from a Dunn's test (M=12s, SD=1.5s) (5.6s difference in mean time, p value = 0.0001). For walking aid recognition there was no significant difference in the accuracies for the amount of layers from a Dunn's test, but there was significant difference in the time it took to process with one layer (M=307.8s, SD=12.43s) being significantly quicker from a Dunn's test (256.6s between 1 and 2 layers and 507.6s between 1 and 3 layers, p<0.0001 when comparing 1 to both 2 and 3 layers). A one-layer LSTM was hence be used for both terrain classification and walking aid recognition. The processing time for the walking aid classification was significantly longer then the terrain classification because the analysis was conducted on a less computationally powerful computer, so can not be used to compare computational time for the two different classifications.

Table 4.7 Percentage accuracies for number of LSTM layers and processing time for terrain classification.

Trial	1 Layer	Processing time (s)	2 Layers	Processing time (s)	3 Layers	Processing time (s)
1	73.91%	7	71.90%	15	67.42%	18
2	73.24%	7	72.62%	12	67.88%	17
3	69.02%	6	72.05%	11	66.91%	18
4	72.88%	6	71.18%	11	63.46%	20
5	72.93%	6	74.68%	11	66.24%	19
Mean	72.39%	6.4	72.49%	12	66.38%	18.4

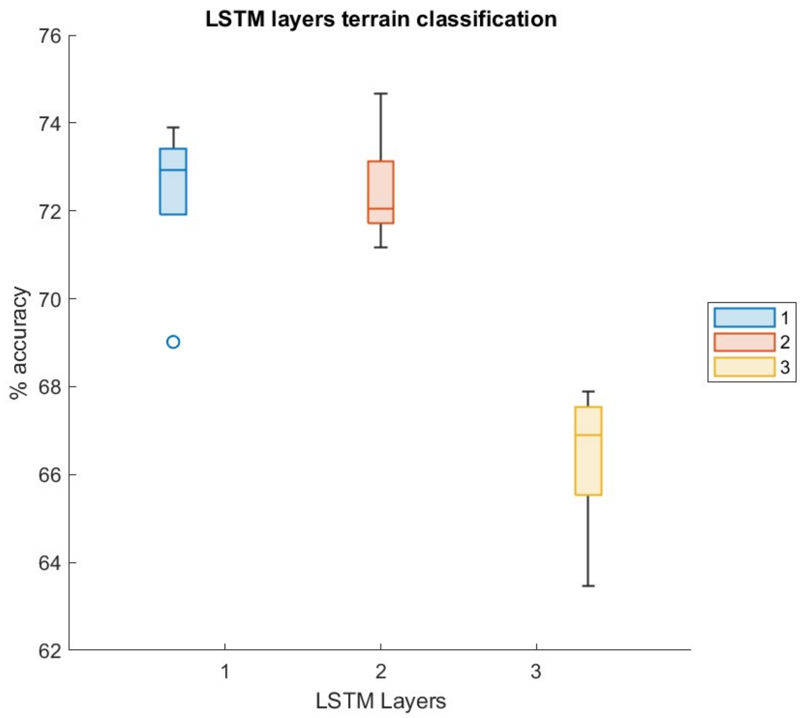

Fig. 4.7 Cross-validation accuracy for terrain classification using LSTM models with 1, 2, and 3 layers. This compares the accuracy of LSTM models with different numbers of layers, indicating that a single LSTM layer achieves the highest accuracy while adding more layers (2 or 3) results in a significant decrease in accuracy. The boxes represent the range of accuracy values across trials, with the median value indicated within each box.

Table 4.8 Percentage accuracies for number of LSTM layers and processing time for walking aid recognition

Trial	1 Layer	Processing time (s)	2 Layers	Processing time (s)	3 Layers	Processing time (s)
1	67.22%	286.00	67.94%	545.00	69.27%	782.00
2	68.62%	313.00	64.97%	560.00	61.01%	796.00
3	67.70%	310.00	69.03%	537.00	69.34%	761.00
4	69.33%	324.00	64.18%	597.00	60.99%	893.00
5	69.74%	306.00	67.42%	548.00	64.49%	845.00
Mean	68.52%	307.80	66.71%	557.40	65.02%	815.40

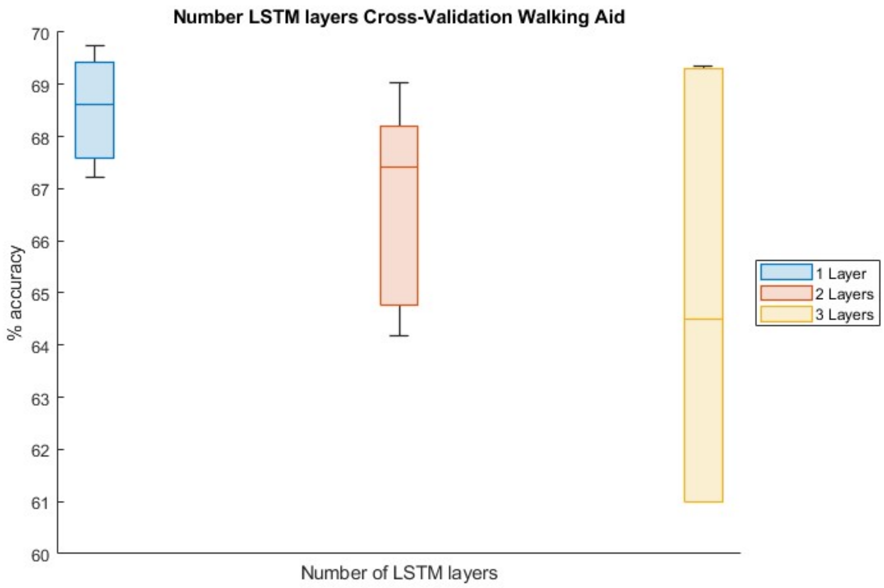

Fig. 4.8 Cross-validation accuracy for terrain classification using LSTM models with 1, 2, and 3 layers. This compares the accuracy of LSTM models with different numbers of layers, indicating that a single LSTM layer achieves the highest accuracy while adding more layers (2 or 3) results in a significant decrease in accuracy. The boxes represent the range of accuracy values across trials, with the median value indicated within each box.

The hyperparameters that were optimised for LSTM were the number of hidden units in the LSTM layer, dropout probability, initial learning rate, minimum batch size, maximum epochs, and validation frequency. For terrain classification the optimised hyperparameters are presented in Table 4.9.

Table 4.9 Hyperparameters for LSTM models

	Number of hidden units	Dropout probability	Initial learn rate	Minimum batch size	Maximum epochs	Validation frequency
Terrain classification	508	0.4	0.027118	127	11	33
Walking aid recognition	693	0.9	0.0096238	95	19	17

Logistic Regression (LR)

A LR classifier was used as a baseline for the other classifiers to compare against, as it is a simple classifier that has been use in previous studies [74]. Due to this, the LR algorithm was not optimised. The LR classifier is only being used for walking aid recognition. This is because LR is used for binary classification so would not be appropriate for the terrain classification.

4.3 Results

There are two parts to the Results: the first part 4.3.2 is concerned with terrain classifications and the second part 4.3.3 with walking aid use classification. For both parts, effects on classification accuracies were explored in relation to use of sensors, windows, data points per window, variables, normalization, prosthesis type and the respective other classification aspect (terrain/walking aid use). The analysis was run sequentially in the order it is documented. This means that the outcome from one analysis was carried forward to the next part of the analysis, for example the prosthetic shank IMU was deemed the most suitable IMU position in section 4.3.2, so only the prosthetic shank IMU data was used for the subsequent analysis. The sequence of the analysis was set to reduce computational time, as the aspects first compared would have a larger possibility to reduce computational time compared to aspects investigated later. For the analysis the null hypothesis was that there as no differences between the variables being compared for each comparison.

4.3.1 Significance tests

When comparing results, Shapiro-Wilk tests were first run to test the data is normally distributed, then the ratio of the standard deviations was calculated to check if the variances were equal. If all the results were deemed to be normally distributed and have equal variance then an ANOVA test was run to determine if there was any significant differences for the classification accuracies within each machine learning model, with the F-value (F), degrees of freedom and p-value (p) reported. If this showed significance, a post-hoc Tukey's HSD tests was run to investigate comparisons between separate variables, with the mean accuracies, standard deviation, mean differences and p-values (p) reported for significant results. If any of the results were not deemed to be normally distributed or had variances in their variances, then a Kruskal–Wallis test was run following the same process as the ANOVA test, with the Chi-square (X^2), degrees of freedom and p value reported. The Chi-square value is reported as an alternative to the H value because the Kruskal-Wallis test statistic (H) follows a Chi-square distribution. If significance was found a Dunn's post-hoc test was run following the same process as the Tukey's HSD test, with the mean accuracies, standard deviation, mean differences and p-values (p) reported. Significance was accepted as p values lower than 0.05. All the statistical analysis was conduced using inbuilt Matlab functions.

4.3.2 Terrain Classification

Effects of sensor location and combination on terrain classification accuracies

Four IMUs were attached to each participant for this study. Terrain classification algorithms were run for the data of each IMU separately, and then for all combined, using 100Hz data split into windows that contained a whole stride split into 100 datapoints, as this did not significantly condense the data and ensured all datasets were consistent. The mean leave-some-out accuracies are displayed in Table 4.10 and visualised in Figure 4.9, the mean cross-validation accuracies are displayed in Table 4.11 and visualised in Figure 4.10. No significance was found between any of the leave-some-out results from Kruskal–Wallis tests (KNN $X^2(4)=8.65$ $p=0.07$, RF $X^2(4)=7.25$ $p=0.12$ and SVM $X^2(4)=7.07$ $p=0.13$, LSTM $X^2(4)=6.59$ $p=0.16$). SVM and RF produce higher accuracies for all IMU locations. For cross-validation results there was significance found between the means for all the algorithms except for LSTM from Kruskal–Wallis tests (KNN $X^2(4)=17.60$ $p=0.0015$, RF $X^2(4)=17.25$ $p=0.0017$ and SVM $X^2(4)=19.62$ $p<0.0001$). The significant results found from the Dunn's post-hoc tests are displayed in Table 4.12 Combining the data from all the IMUs produced significantly higher accuracies with SVM and RF algorithms to the trunk and other shank IMUs. The prosthetic shank and thigh IMU were significantly better than the trunk IMU for KNN, the thigh IMU also produced significantly higher accuracies than the other shank IMU for KNN. Overall, these results disprove the null hypothesis, that IMU location and combining IMU does not have an effect on terrain classification accuracy.
The leave-some-out accuracies are low, but these are comparable to the only other study that has investigated comparable real-world terrain classification for lower limb prosthetic users (Jamieson, 2021).

As the ideal system will be prosthetic mounted and there is not a significant difference between using all IMU to just the prosthetic shank IMU in leave-some-out verification, and only a slight decrease in cross-validation, only the prosthetic shank IMU was used in the next stages of analysis.

Table 4.10 Leave-some-out terrain classification percentage accuracies for different IMU positions.

	KNN	RF	SVM	LSTM
Prosthetic Shank	44.19%	52.09%	52.87%	47.53%
Other Shank	43.52%	53.56%	52.41%	46.62%
Thigh	37.53%	44.36%	43.95%	39.64%
Trunk	43.51%	54.3%	53.22%	50.31%
All	44.73%	52.00%	49.08%	50.06%

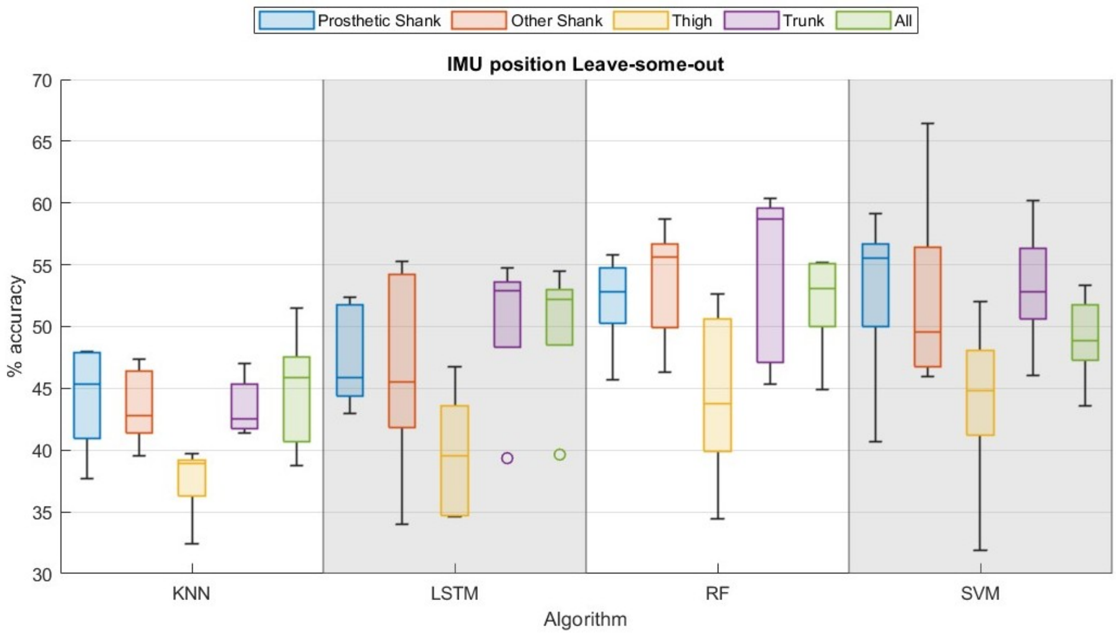

Fig. 4.9 Leave-some-out accuracy results for terrain classification across different sensor placements and algorithms. The boxplot shows the classification accuracy for each sensor placement (prosthetic shank, other shank, thigh, trunk, and all combined) using four machine learning algorithms: K-Nearest Neighbors (KNN), Long Short-Term Memory (LSTM), Random Forest (RF), and Support Vector Machines (SVM). The median and inter-quartile range of the accuracy for each sensor placement and algorithm is indicated by the spread of the boxes and the maximum and minimum values, if these are not outliers, are presented by the whiskers.

Table 4.11 Cross-validation terrain classification percentage accuracies for different IMU positions.

	KNN	RF	SVM	LSTM
Prosthetic Shank	84.30%	81.66%	84.29%	81.99%
Other Shank	83.32%	80.60%	82.11%	82.20%
Thigh	84.72%	82.11%	83.15%	83.21%
Trunk	81.67%	79.90%	81.36%	81.48%
All	82.68%	84.92%	87.32%	81.61 %

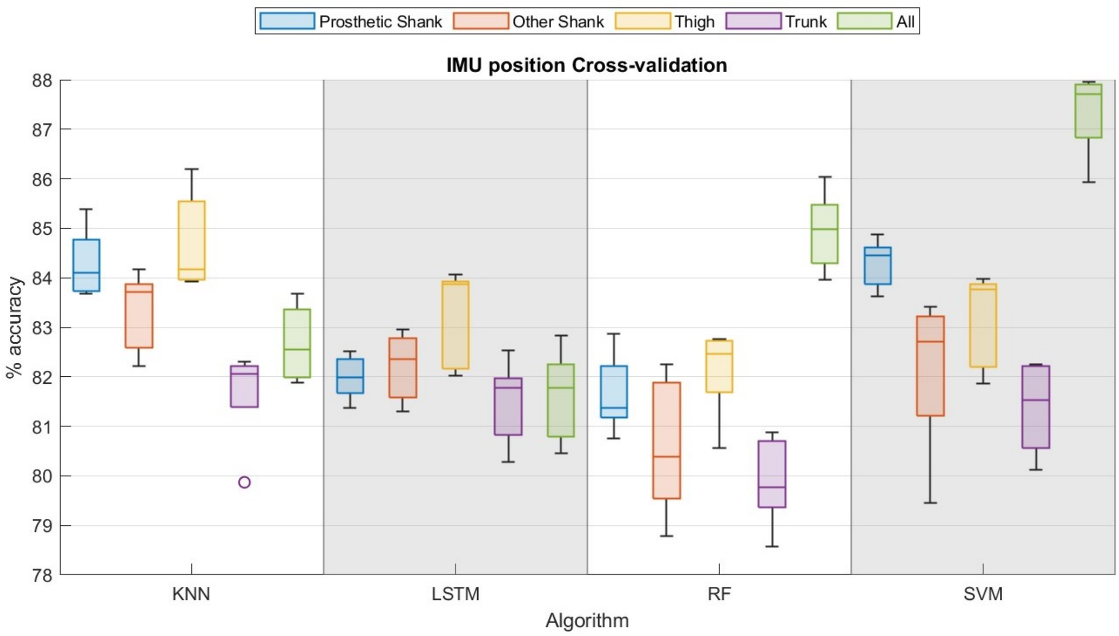

Fig. 4.10 Cross-validation accuracy results for terrain classification across different sensor placements and algorithms. The boxplot shows the classification accuracy for each sensor placement (prosthetic shank, other shank, thigh, trunk, and all combined) using four machine learning algorithms: K-Nearest Neighbors (KNN), Long Short-Term Memory (LSTM), Random Forest (RF), and Support Vector Machines (SVM). Combined sensors show higher accuracy, particularly when using RF and SVM algorithms. The prosthetic shank IMU placement showed higher accuracies than the other placements when using SVM algorithm. The median and inter-quartile range of the accuracy for each sensor placement and algorithm is indicated by the spread of the boxes and the maximum and minimum values, if these are not outliers, are presented by the whiskers.

Table 4.12 Significant differences in mean accuracies for cross-validation test between IMU placements and algorithms.

Model		Mean (%) ± *SD*		Mean (%) ± *SD*	Difference (%)	*p*-value
SVM	All	87.32 ± 0.76	Trunk	81.36 ± 0.84	5.96	0.0009
	All	87.32 ± 0.76	OS	82.11 ± 1.44	5.21	0.011
RF	All	84.93 ± 0.72	Trunk	79.90 ± 0.82	5.02	0.001
	All	84.93 ± 0.72	OS	80.60 ± 1.27	4.62	0.017
KNN	PS	84.30 ± 0.63	Trunk	81.67 ± 0.91	2.63	0.026
	Thigh	84.72 ± 0.90	Trunk	81.67 ± 0.91	3.05	0.005
	Thigh	84.72 ± 0.90	OS	83.33 ± 0.73	1.40	0.005

Effects of window type on terrain classification accuracies

To assess how different window types affect the accuracies of terrain classification eight different window types were assessed: six time-windows (0.1s, 0.2s, 0.5s, 1s, 1.5s, 2s), a trial where the window was made to cover one stride and assessing individual datapoints. For all these window types, accuracy was quantified as the percentage of strides correctly classified. For time-windows smaller than a stride the median classified terrain over the multiple windows that made up the stride was used. For windows that are larger than a stride, the terrain that was the most predominant across the window was used as the terrain for the whole window. Tables 4.13 and 4.14 display the mean accuracies for leave-some-out and cross-validation, with respective results visualised in Figures 4.11 and 4.12. Significance was found between any of the leave-some-out results from Kruskal–Wallis tests (KNN $X^2(7)$=25.34 p<0.001, RF $X^2(7)$=16.32 p=0.022 and SVM $X^2(7)$=19.08 p=0.008, LSTM $X^2(7)$=20.39 p=0.0048). The individual datapoints were not validated using cross-validation as the algorithms would have been trained on datapoints of the same stride and would produce artificially high accuracies. In the leave-one-out tests the stride-based window produced significantly higher accuracies than the 2s window for RF and the individual datapoints for SVM. The only other significant result for leave-some-out tests was the 2s window producing lower accuracies than 0.1s window for KNN and LSTM. For cross-validation verification, significance was found between any of the leave-some-out results from Kruskal–Wallis tests (KNN $X^2(6)$=33.33 p<0.001, RF $X^2(6)$=33.24 p<0.001 and SVM $X^2(6)$=32.69 p<0.001, LSTM $X^2(6)$=32.73 p<0.001). Dunn's post-hoc tests found the stride-based window produced significantly higher accuracies than 2s window for every algorithm, 1.5s window for KNN and RF, 1s window for KNN, 0.2s window for SVM and LSTM and 0.1s window for SVM. Results for significance testing are displayed in Table 4.15. These results disprove the null hypothesis and show that the windowing method does have an effect on terrain classification accuracy. The stride-based windows produced the best accuracies for both leave-some-out and cross-validation, so they were the windows that were used for the rest of the analysis.

Table 4.13 Leave-some-out terrain classification percentage accuracies for different window types.

	KNN	RF	SVM	LSTM
0.1s	43.97%	43.39%	37.46%	48.05%
0.2s	39.99%	43.04%	41.56%	44.31%
0.5s	39.99%	43.11%	41.56%	44.46%
1s	36.49%	43.01%	43.24%	39.40%
1.5s	39.72%	42.88%	39.09%	38.60%
2s	36.35%	42.11%	37.53%	37.03%
Stride	44.19%	52.09%	52.87%	47.53%
Individual	44.42%	46.52%	32.72%	48.23%

Table 4.14 Cross-validation terrain classification percentage accuracies for different window types.

	KNN	RF	SVM	LSTM
0.1s	83.40%	79.70%	48.58%	76.75%
0.2s	77.68%	74.32%	51.50%	54.94%
0.5s	68.26%	69.53%	61.59%	69.32%
1s	63.01%	62.79%	64.54%	67.15%
1.5s	60.28%	60.85%	63.98%	67.00%
2s	56.70%	56.08%	57.72%	61.93%
Stride	84.30%	81.66%	84.29%	81.99%

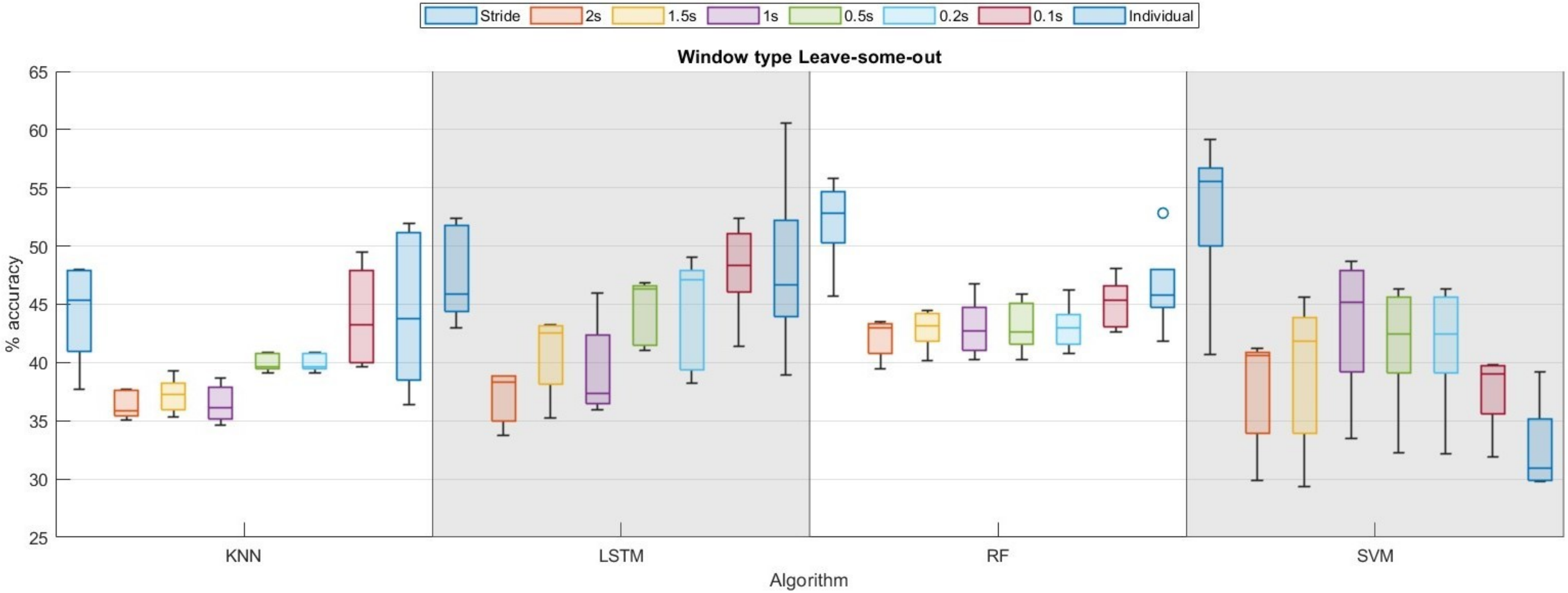

Fig. 4.11 Leave-some-out accuracy results for terrain classification across different windowing methods. The boxplot shows the classification accuracy for windowing method (time-based windows 0.1s, 0.2s, 0.5s, 1s, 1.5s and 2s, stride-based window and individual datapoints) using four machine learning algorithms: K-Nearest Neighbors (KNN), Long Short-Term Memory (LSTM), Random Forest (RF), and Support Vector Machines (SVM). The stride-based window produced significantly higher accuracies than the 2s window for RF and the individual datapoints for SVM. The 0.1s window produced higher accuracies than the 2s window for KNN and LSTM.

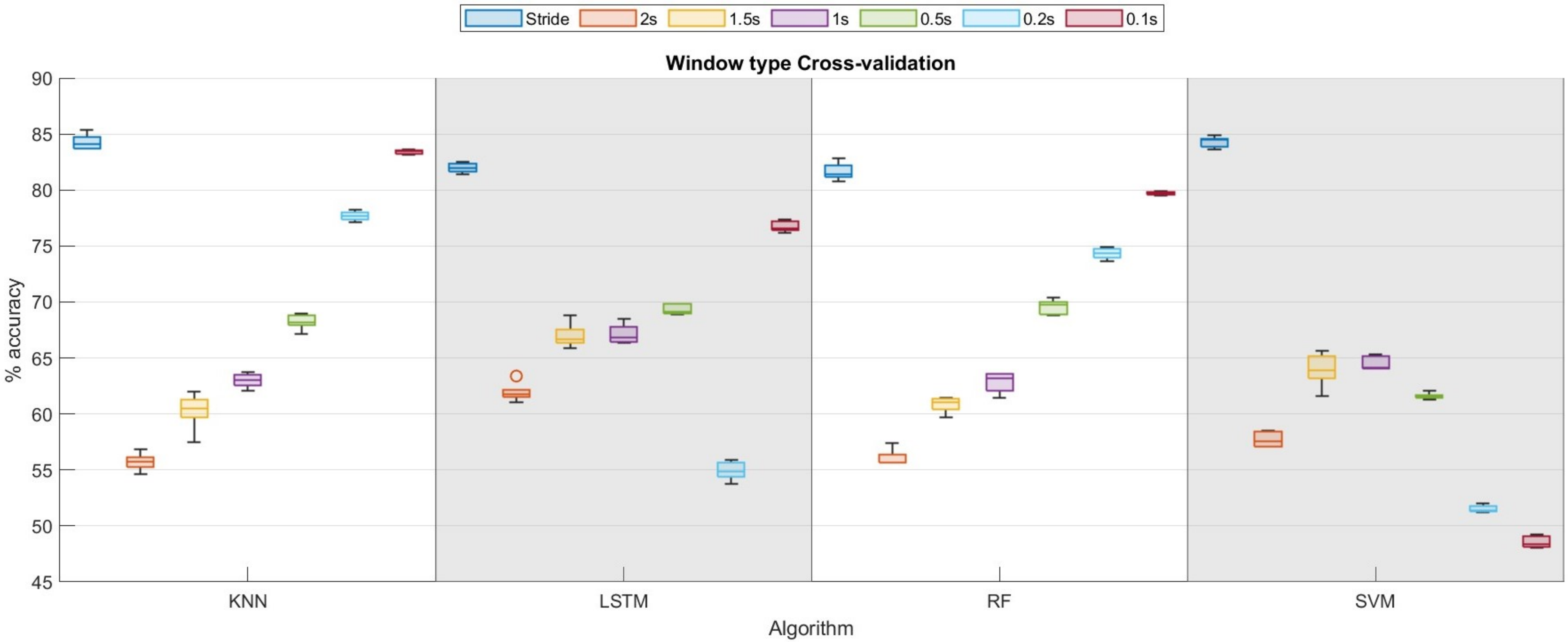

Fig. 4.12 Cross-validation accuracy results for terrain classification across different windowing methods. The boxplot shows the classification accuracy for windowing method (time-based windows 0.1s, 0.2s, 0.5s, 1s, 1.5s and 2s, stride-based window and individual datapoints) using four machine learning algorithms: K-Nearest Neighbors (KNN), Long Short-Term Memory (LSTM), Random Forest (RF), and Support Vector Machines (SVM). The stride-based window produced significantly higher accuracies than all the other windows for at least one classifier.

Table 4.15 Significant differences in mean accuracies for terrain classification across different window types. ID = Individual Datapoints

Model		Mean (%) ± *SD*		Mean (%) ± *SD*	Mean Difference (%)	*p*-value
Leave-Some-Out Results						
RF	Stride	52.09 ± 3.48	2s	42.11 ± 1.53	9.98	0.043
SVM	Stride	52.87 ± 6.40	ID	32.72 ± 3.53	20.15	0.006
KNN	0.1s	43.97 ± 3.92	2s	36.35 ± 1.10	7.62	0.046
LSTM	0.1s	48.05 ± 3.73	2s	37.03 ± 2.09	11.02	0.022
Cross-Validation Results						
KNN	Stride	84.30 ± 0.63	2s	55.70 ± 0.72	27.6	<0.001
	Stride	84.30 ± 0.63	1.5s	60.28 ± 1.53	24.02	0.002
	Stride	84.30 ± 0.63	1.5s	60.28 ± 1.53	24.02	0.002
	Stride	84.30 ± 0.63	1s	63.01 ± 0.58	21.29	0.042
RF	Stride	81.66 ± 0.72	2s	56.08 ± 0.67	25.58	<0.001
	Stride	81.66 ± 0.72	1.5s	60.85 ± 0.63	20.81	0.003
SVM	Stride	84.29 ± 0.44	2s	57.72 ± 0.63	26.57	0.042
	Stride	84.29 ± 0.44	0.2s	51.50 ± 0.30	32.79	0.002
	Stride	84.29 ± 0.44	0.1s	84.58 ± 0.47	35.71	<0.001
LSTM	Stride	81.99 ± 0.40	2s	61.93 ± 0.77	20.06	0.002
	Stride	81.99 ± 0.40	0.2s	54.94 ± 0.74	27.05	<0.001

Effect of datapoints per window on terrain classification accuracies.

The more data points that make up a stride, the more computational time it will take to process the data, but using fewer data points reduces the amount of information and could reduce the accuracy of classifiers. Four different splits of stride windows were assessed (10, 20, 50 and 100). Table 4.16 and 4.17 display the mean accuracies for leave-some-out and cross-validation for the four window datapoint numbers, and Figures 4.13 and 4.14 visualise these results. There was no statistical difference between the number of datapoints used, and results were statistically similar for all algorithms from Kruskal–Wallis tests (KNN $X^2(3)=0.40$ $p=0.940$, RF $X^2(3)=1.39$ $p=0.708$, SVM $X^2(3)=0.34$ $p=0.953$ and LSTM $X^2(3)=0.25$ $p=0.970$) for leave-some-out verification. For cross-validation, there was no statistical significance between RF and LSTM for the four datapoint quantities, but KNN and SVM showed significant differences from Kruskal–Wallis tests (KNN $X^2(3)$=8.47 p=0.0372, RF $X^2(3)$=4.56 p=0.21, SVM $X^2(3)$=17.86 p<0.001 and LSTM $X^2(3)$=6.34 p=0.096). Running Dunn's post-hoc tests found no difference for the datapoint amounts for KNN but, as shown in Figure 4.14, there is clear significant difference for SVM, where less datapoints produced lower accuracies. Overall, these results do not disprove the null hypothesis that the number of datapoints in a window effects terrain classification accuracies. As there is only a difference for cross-verification SVM, 10 datapoint stride windows were used for the rest of the analysis, but the final algorithm was assessed at 100 datapoints as well.

Table 4.16 Leave-some-out terrain classification percentage accuracies for different numbers of datapoints per window

	KNN	RF	SVM	LSTM
10	45.45%	54.40%	52.10%	45.58%
20	44.55%	53.13%	51.78%	47.35%
50	44.31%	53.25%	52.10%	47.39%
100	44.19%	52.09%	52.87%	47.53%

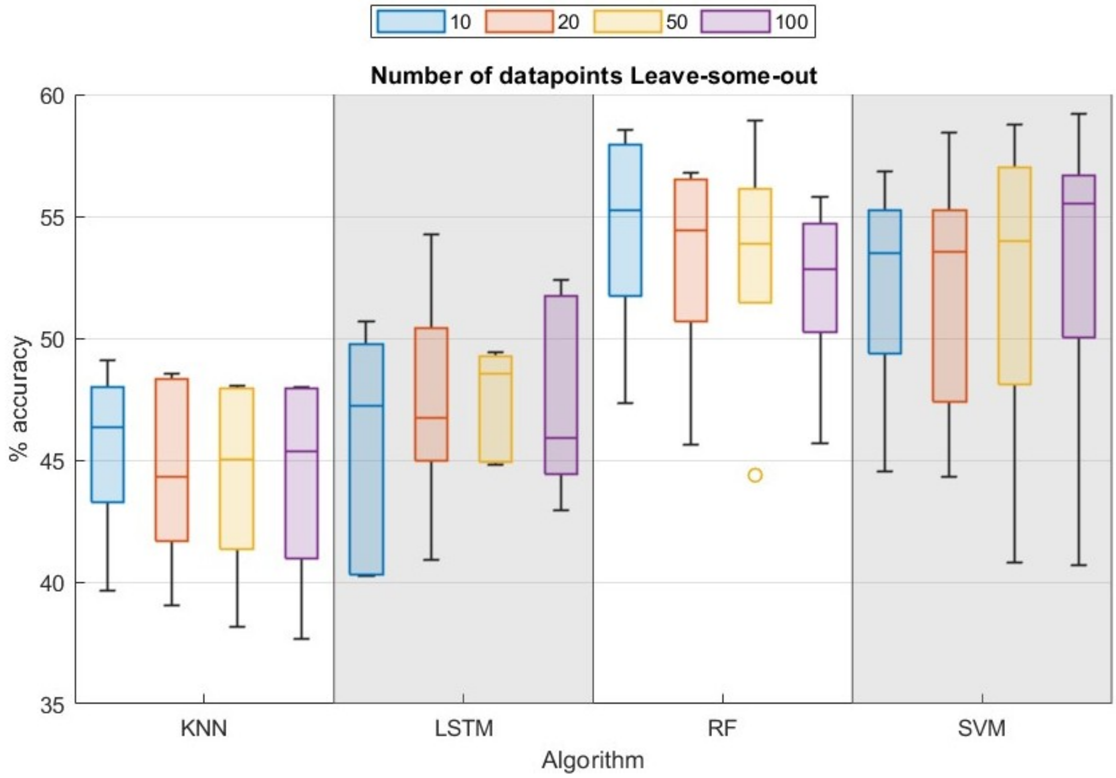

Fig. 4.13 Leave-some-out accuracy results for terrain classification across different number of datapoints per window. The boxplot shows the classification accuracy for each number of datapoints (10, 20, 50, 100) using four machine learning algorithms: K-Nearest Neighbors (KNN), Long Short-Term Memory (LSTM), Random Forest (RF), and Support Vector Machines (SVM). The number of datapoints didn't affect the accuracies using all algorithms.

Table 4.17 Cross-validation terrain classification percentage accuracies for different numbers of datapoints per window

	KNN	RF	SVM	LSTM
10	85.27%	82.59%	77.31%	80.78%
20	85.02%	82.32%	79.95%	81.65%
50	84.34%	81.62%	82.10%	82.23%
100	84.30%	81.66%	84.29%	81.99%

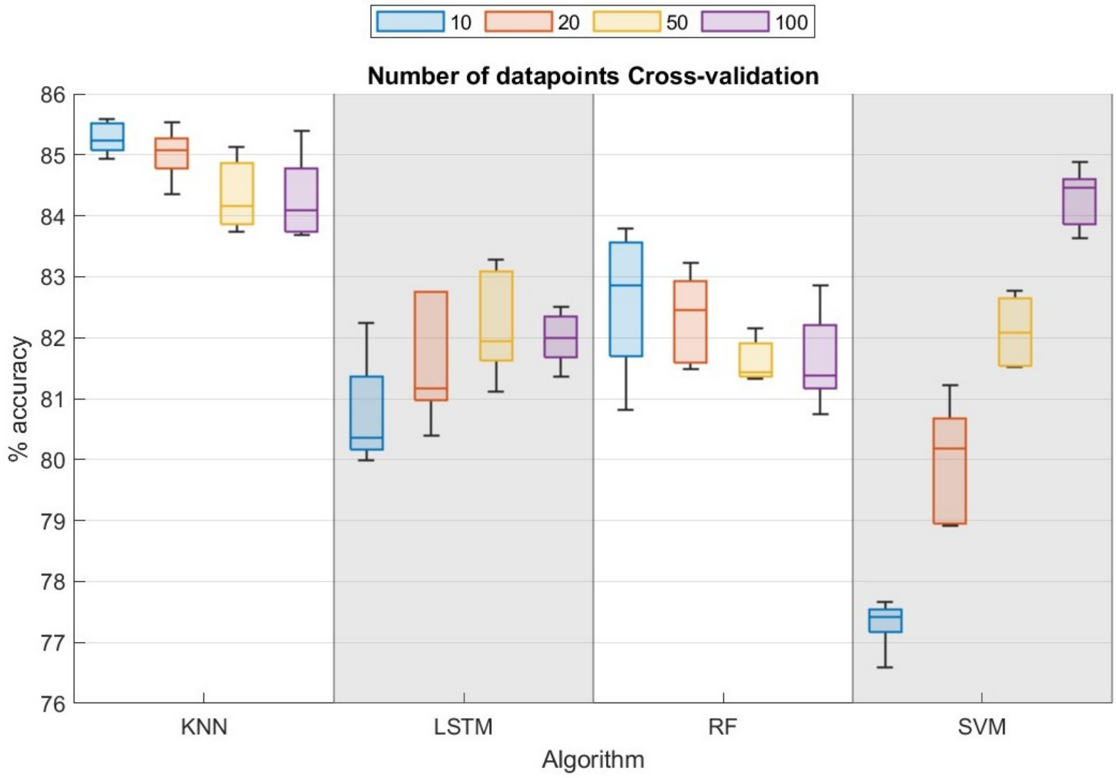

Fig. 4.14 Cross-validation accuracy results for terrain classification across different number of datapoints per window. The boxplot shows the classification accuracy for each number of datapoints (10, 20, 50, 100) using four machine learning algorithms: K-Nearest Neighbors (KNN), Long Short-Term Memory (LSTM), Random Forest (RF), and Support Vector Machines (SVM). The only significant difference with using fewer datapoints was with the SVM algorithm, the number of datapoints didn't affect the accuracies using the other algorithms.

Effects of variables used on terrain classification accuracies.

Reducing the number of features reduces the computational burden associated with running a classifier. Also, understanding which IMU sensors and associated features are critical to the accuracies of the classifiers could allow for a simpler system or reduced system requirements. For this analysis the variables compared were:

- Accelerations – X, Y, and Z components of the acceleration and the resultant acceleration
- Gyroscope - X, Y, and Z components of the gyroscopic data
- Magnetometer - X, Y, and Z components of the magnetometer data
- Free accelerations - X, Y, and Z components of the free acceleration and the resultant free acceleration

- Cadence – the cadence per stride
- Velocity - X, Y, and Z components of the velocity data
- AGM – Acceleration, Gyroscope and magnetometer data
- AGMF - Acceleration, Gyroscope, magnetometer and free acceleration data
- AGMFC - Acceleration, Gyroscope, magnetometer, free acceleration and cadence data
- All – all the data

The mean accuracies for leave-some-out and cross-validation verification are displayed in Tables 4.18 and 4.19, and the results visualised in Figures 4.15 and 4.16. Significance tests showed significant differences for all the algorithms from Kruskal–Wallis tests (KNN $X^2(9)$=40.66 p<0.0001, RF $X^2(9)$=23.70 p<0.0001, SVM $X^2(9)$=42.23 p<0.0001, LSTM $X^2(9)$=31.42 p<0.0001) for leave-some-out validation, which looking at the results is expected because there is clear difference between the box plots for cadence and the other variables for each algorithm. Dunn's post-hoc tests found the gyroscope produced significantly lower accuracies than AGMFC and all for KNN, RF and SVM, and lower than AGMF for RF and SVM. Cadence produced significantly lower accuracies than AGMFC and all for all algorithms, AGMF for KNN, RF and SVM and AGM for SVM. There was no statistical significance between the combined variables AGM, AGMF, AGMFC and 'All'. For the individual variables the only significant difference was obtained for free acceleration producing significantly higher accuracies than cadence for LSTM. For the cross-validation assessment, significance tests showed significant differences for all the algorithms from Kruskal–Wallis tests (KNN $X^2(9)$=45.41 p<0.0001, RF $X^2(9)$=44.75 p<0.0001, SVM $X^2(9)$=47.99 p<0.0001, LSTM $X^2(9)$=45.28p<0.0001). Dunn's post-hoc tests found cadence came out significantly worse than all combined features. Gyroscope data produced significantly lower accuracies than AGMFC for all algorithms, all the data for RF, SVM and LSTM, AGMF for KNN, RF and LSTM and AGM for RF and LSTM. Magnetometer accuracies were significantly lower than all the data for SVM and LSTM, and AGMFC for SVM. The velocity data accuracies were lower than all the data for SVM. The post-hoc tests did not show a significant difference between any of the individual variables, but cadence produced lower mean accuracies. There was no significant difference between the combined features, but AGM produce lower mean accuracies than other combined variables. Results for significance testing are displayed in Table 4.20. These results disprove the null hypothesis and show that the variables used to in terrain classification do have an effect on the classification accuracies.

Since AGMFC produced higher mean accuracies than the other combinations on 3 of the leave-some-out tests, it will therefore be used for the rest of the analysis.

Table 4.18 Leave-some-out terrain classification percentage accuracies for different variables. AGM = accelerations, gyroscope and magnetometer, AGMF = accelerations, gyroscope, magnetometer and free accelerations, AGMFC= accelerations, gyroscope, magnetometer, free accelerations and cadence.

	KNN	RF	SVM	LSTM
Accelerations	36.80%	44.12%	42.14%	40.36%
Gyroscope	29.47%	36.32%	29.03%	32.86%
Magnetometer	28.57%	41.48%	35.80%	34.00%
Free Accelerations	40.91%	50.03%	48.60%	46.22%
Cadence	21.85%	26.79%	23.37%	28.53%
Velocity	34.50%	41.82%	37.43%	40.08%
AGM	41.82%	49.49%	49.39%	41.15%
AGMF	44.87%	54.18%	52.28%	43.32%
AGMFC	45.46%	54.36%	52.83%	46.94%
All	45.45%	54.40%	52.10%	45.58%

Table 4.19 Cross-validation terrain classification percentage accuracies for different variables. AGM = accelerations, gyroscope and magnetometer, AGMF = accelerations, gyroscope, magnetometer and free accelerations, AGMFC= accelerations, gyroscope, magnetometer, free accelerations and cadence.

	KNN	RF	SVM	LSTM
Accelerations	76.46%	76.26%	58.96%	71.60%
Gyroscope	73.66%	74.07%	48.26%	62.75%
Magnetometer	80.77%	79.47%	47.98%	66.74%
Free Accelerations	80.04%	77.52%	57.54%	74.16%
Cadence	33.65%	36.29%	25.75%	30.45%
Velocity	76.07%	76.21%	54.06%	72.64%
AGM	84.85%	82.68%	70.08%	79.49%
AGMF	85.67%	83.18%	74.44%	80.18%
AGMFC	85.61%	82.83%	75.17%	80.59%
All	85.27%	82.59%	77.31%	80.78%

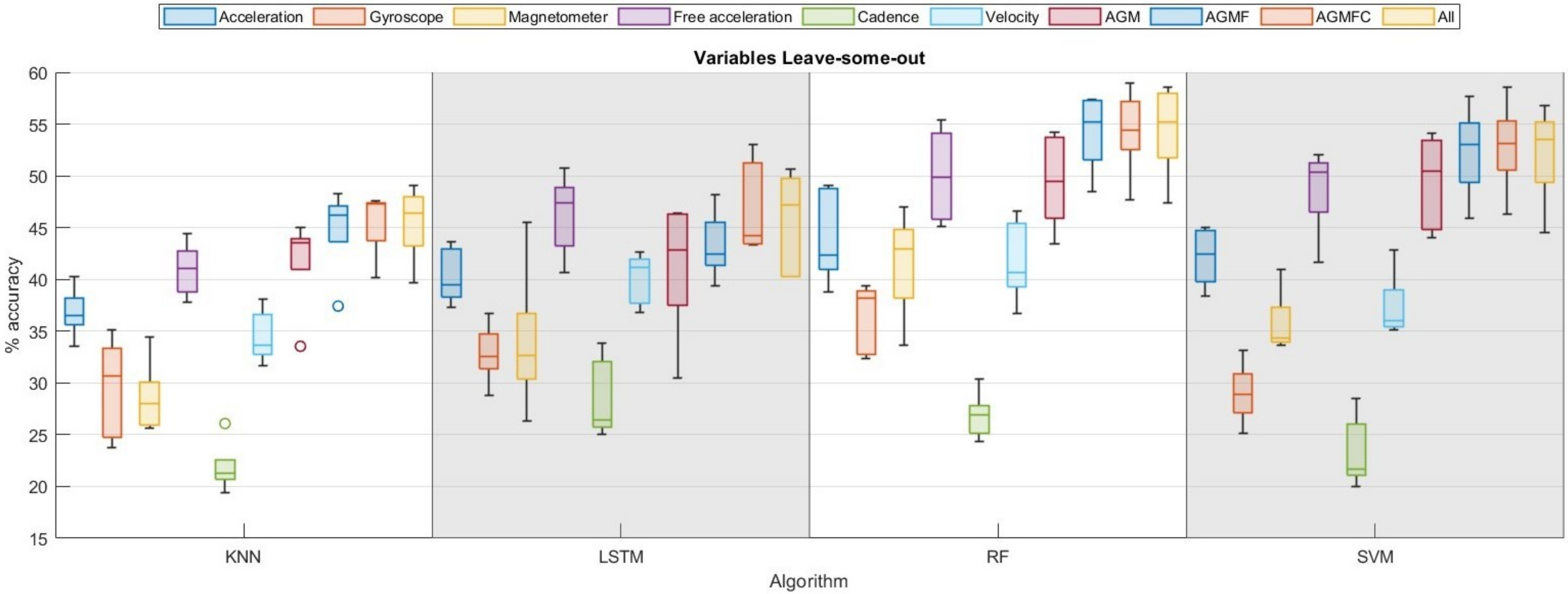

Fig. 4.15 Leave-some-out accuracy results for terrain classification across different variables and variable groups. The boxplot shows the classification accuracy for each variable (acceleration, gyroscope, magnetometer, free accelerations, cadence, velocity, AGM = accelerations, gyroscope and magnetometer, AGMF = accelerations, gyroscope, magnetometer and free accelerations, AGMFC= accelerations, gyroscope, magnetometer, free accelerations and cadence.) using four machine learning algorithms: K-Nearest Neighbors (KNN), Long Short-Term Memory (LSTM), Random Forest (RF), and Support Vector Machines (SVM). The combined variable groups produced higher accuracies than the individual variables. AGMFC produced higher mean accuracies than the other combinations on 3 of the leave-some-out tests.

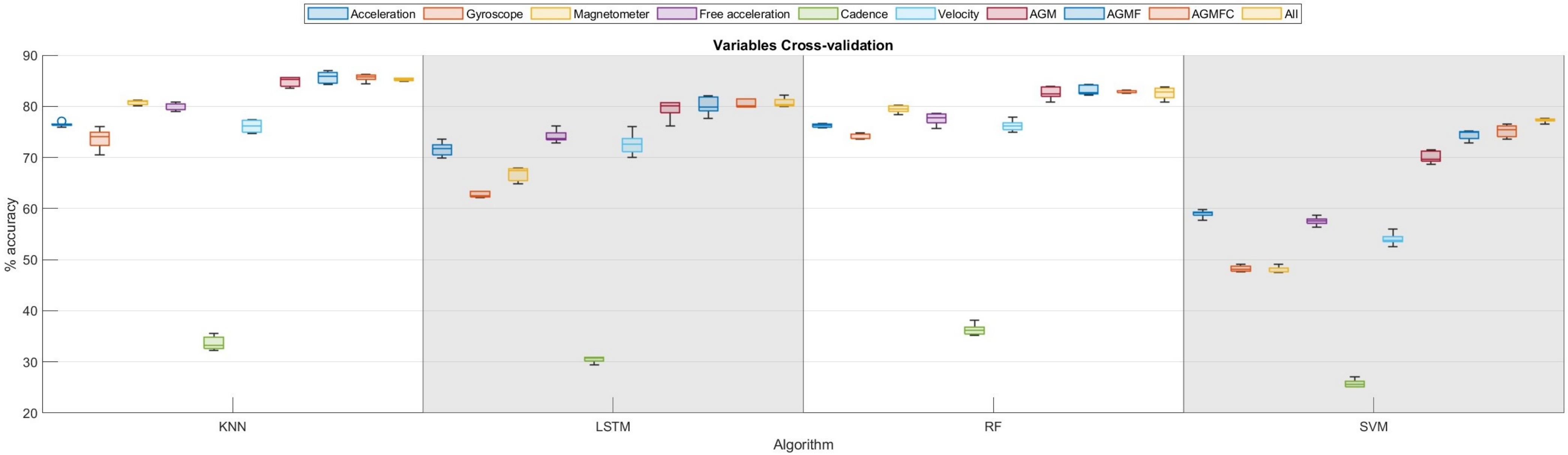

Fig. 4.16 Cross-validation accuracy results for terrain classification across different variables and variable groups. The boxplot shows the classification accuracy for each variable (acceleration, gyroscope, magnetometer, free accelerations, cadence, velocity, AGM = accelerations, gyroscope and magnetometer, AGMF = accelerations, gyroscope, magnetometer and free accelerations, AGMFC= accelerations, gyroscope, magnetometer, free accelerations and cadence.) using four machine learning algorithms: K-Nearest Neighbors (KNN), Long Short-Term Memory (LSTM), Random Forest (RF), and Support Vector Machines (SVM). The combined variable groups produced higher accuracies than the individual variables.

Table 4.20 Significant differences in mean accuracies for terrain classification across different variables. Gyr = Gyroscope, Cad = Cadence, Mag = Magnetometer, Vel = Velocity

Model		Mean (%) ± *SD*		Mean (%) ± *SD*	Difference (%)	*p*-value
Leave-Some-Out Results						
KNN	Gyro	29.47 ± 4.89	AGMFC	45.46 ± 2.65	15.99	0.05
	Gyro	29.47 ± 4.89	All	45.45 ± 3.28	15.98	0.046
	Cad	21.85 ± 3.48	AGMFC	45.46 ± 2.65	23.61	0.002
	Cad	21.85 ± 3.48	All	45.45 ± 3.28	23.6	0.002
	Cad	21.85 ± 3.48	AGMF	44.87 ± 3.82	23.02	0.004
RF	Gyro	36.32 ± 3.03	AGMFC	54.36 ± 3.79	18.04	0.029
	Gyro	36.32 ± 3.03	All	54.40 ± 4.01	18.08	0.029
	Gyro	36.32 ± 3.03	AGMF	54.18 ± 3.35	17.89	0.04
	Cad	26.79 ± 2.02	AGMFC	54.36 ± 3.79	27.57	0.002
	Cad	26.79 ± 2.02	All	54.40 ± 4.01	27.61	0.002
	Cad	26.79 ± 2.02	AGMF	54.18 ± 3.35	27.39	0.003
SVM	Gyro	29.03 ± 2.65	AGMFC	52.82 ± 3.98	23.8	0.011
	Gyro	29.03 ± 2.65	All	52.10 ± 4.23	23.07	0.017
	Gyro	29.03 ± 2.65	AGMF	52.28 ± 3.94	23.25	0.018
	Cad	23.37 ± 3.08	AGMFC	52.82 ± 3.98	29.46	0.002
	Cad	23.37 ± 3.08	All	52.10 ± 4.23	28.73	0.003
	Cad	23.37 ± 3.08	AGMF	52.28 ± 3.94	28.91	0.004
	Cad	23.37 ± 3.08	AGM	49.39 ± 4.12	22.7	0.034
LSTM	Cad	28.53 ± 3.48	AGMFC	46.94 ± 4.09	18.41	0.007
	Cad	28.53 ± 3.48	All	45.58 ± 4.47	17.05	0.032
	Cad	28.53 ± 3.48	FA	46.94 ± 4.09	13.36	0.009
Cross-Validation						
KNN	Gyro	73.66 ± 1.84	AGMFC	85.61 ± 0.63	11.95	0.009
	Gyro	73.66 ± 1.84	AGMF	85.67 ± 1.04	12.01	0.011
RF	Gyro	74.07 ± 0.50	AGMFC	82.83 ± 0.25	8.76	0.018
	Gyro	74.07 ± 0.50	All	82.59 ± 1.08	8.52	0.028
	Gyro	74.07 ± 0.50	AGMF	83.18 ± 0.85	9.11	0.009
	Gyro	74.07 ± 0.50	AGM	84.85 ± 0.85	8.61	0.028
SVM	Gyro	48.26 ± 0.53	AGMFC	75.17 ± 1.09	26.91	0.046
	Gyro	48.26 ± 0.53	All	77.31 ± 0.37	29.05	0.032
	Mag	47.98 ± 0.58	All	77.31 ± 0.37	29.33	0.001
	Mag	47.98 ± 0.58	AGMFC	75.17 ± 1.09	27.19	0.023
	Vel	54.06 ± 1.10	All	77.31 ± 0.37	23.25	0.05
LSTM	Gyro	62.75 ± 0.56	AGMFC	80.59 ± 0.75	17.84	0.017
	Gyro	62.75 ± 0.56	All	80.78 ± 0.82	18.03	0.006
	Gyro	62.75 ± 0.56	AGMF	80.18 ± 1.60	17.43	0.027
	Gyro	62.75 ± 0.56	AGM	79.49 ± 1.73	16.74	0.043
	Mag	66.74 ± 1.25	All	80.78 ± 0.82	14.04	0.046

Effects of normalisation on terrain classification.

The data were normalised against the mean for the individual person and individual strides. Leave-some-out and cross-validation mean accuracies are displayed in Tables 4.21 and 4.22, with the results visualised in Figures 4.17 and 4.18. For all trials, normalising per stride produced smaller accuracies, so a combined trial was run with raw and normalised per person data (RP). For leave-some-out there was significant differences found only between RF and LSTM results from Kruskal–Wallis tests (KNN $X^2(4)=6.71$ $p=0.15$, RF $X^2(4)=9.56$ $p=0.048$, SVM $X^2(4)=5.00$ $p=0.29$, LSTM $X^2(4)=9.59$ $p=0.048$), but post-hoc test found no significant difference between individual normalisation methods. For cross-validation, there was significance found in all models from Kruskal–Wallis tests (KNN $X^2(4)=18.07$ $p=0.0012$, RF $X^2(4)=19.37$ $p<0.001$, SVM $X^2(4)=22.90$ $p<0.001$, LSTM $X^2(4)=17.21$ $p=0.002$). From Dunn's post-hoc tests normalised per stride produced significantly lower accuracies than the raw data for RF, normalised per person, RP and all the data for LSTM and RP and all the data for SVM. The raw data were significantly lower than all the data for SVM, but these accuracies were not as high as the other algorithms. Results for significance testing are displayed in Table 4.23. These results disprove the null hypothesis and show that normalisation method does have an effect on terrain classification accuracies. RP data will be used for the rest of the analysis, as it produced the highest mean accuracies for 4 of the tests.

Table 4.21 Leave-some-out terrain classification percentage accuracies for different normalisation techniques. RP = raw and person combined.

	KNN	RF	SVM	LSTM
Raw	42.22%	51.40%	51.55%	46.03%
Stride	37.73%	45.37%	46.85%	41.89%
Person	44.99%	53.79%	52.00%	48.17%
RP	45.67%	53.81%	51.39%	48.77%
All	45.46%	54.36%	52.83%	46.94%

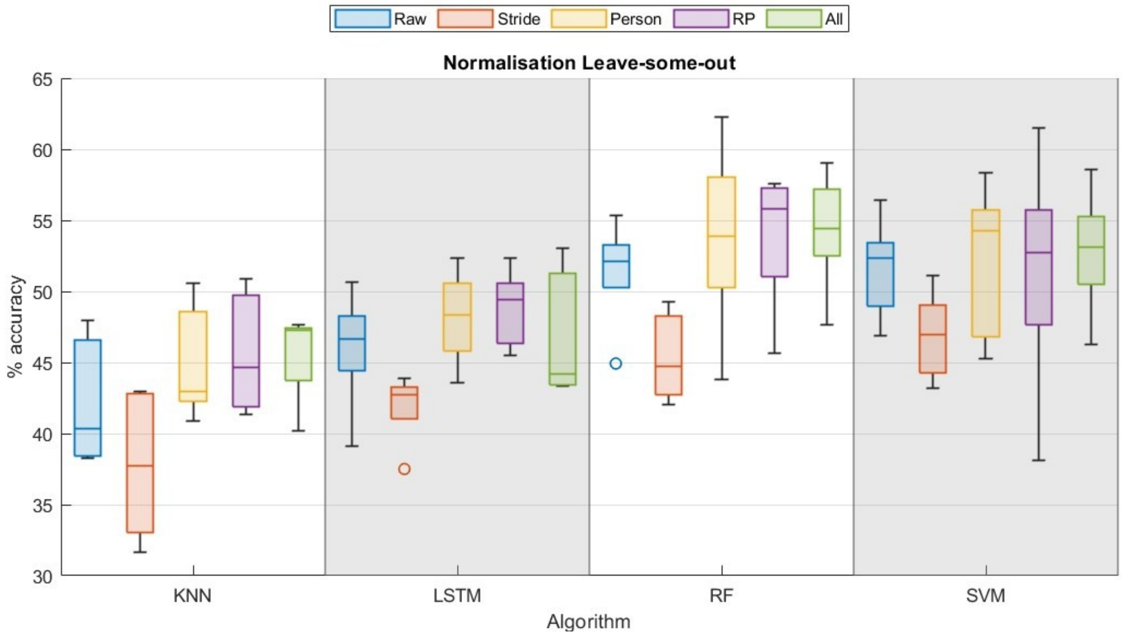

Fig. 4.17 Leave-some-out accuracy results for terrain classification across different normalisation methods. The boxplot shows the classification accuracy for each variable (raw data, normalised per stride, normalised per person, RP = raw and person combined and all data) using four machine learning algorithms: K-Nearest Neighbors (KNN), Long Short-Term Memory (LSTM), Random Forest (RF), and Support Vector Machines (SVM). There was no significant difference between the normalisation techniques.

Table 4.22 Cross-validation terrain classification percentage accuracies for different normalisation techniques. RP = raw and person combined.

	KNN	RF	SVM	LSTM
Raw	83.03%	84.91%	65.62%	77.90%
Stride	82.00%	78.69%	61.93%	76.53%
Person	85.24%	82.12%	66.57%	80.99%
RP	85.71%	83.00%	71.36%	81.00%
All	85.61%	82.83%	75.17%	80.59%

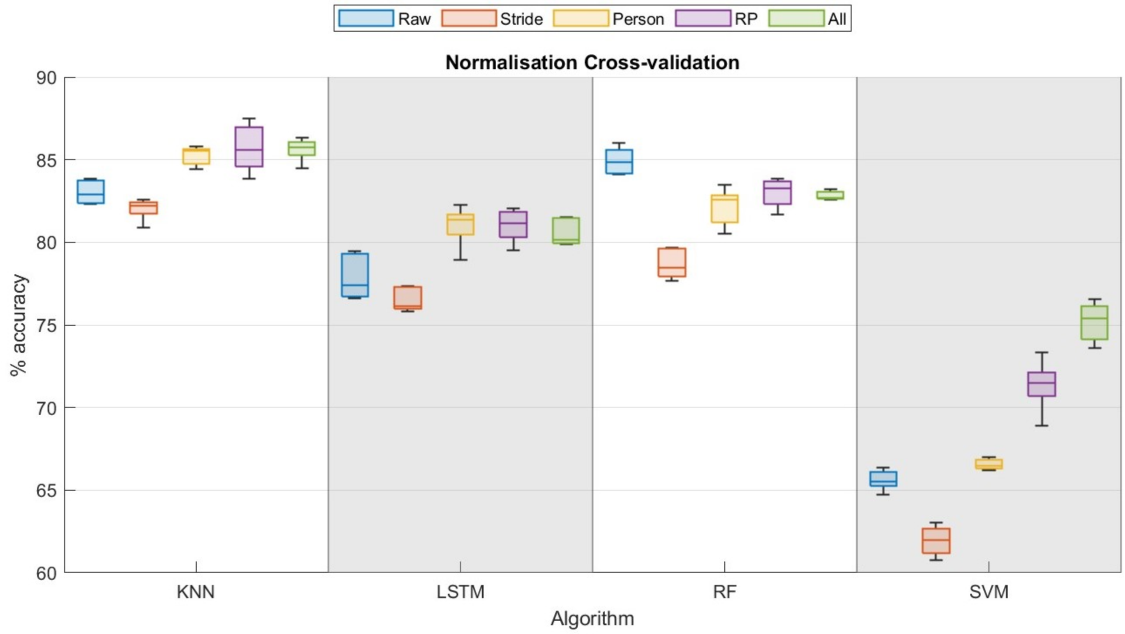

Fig. 4.18 Cross-validation accuracy results for terrain classification across different normalisation methods. The boxplot shows the classification accuracy for each variable (raw data, normalised per stride, normalised per person, RP = raw and person combined and all data) using four machine learning algorithms: K-Nearest Neighbors (KNN), Long Short-Term Memory (LSTM), Random Forest (RF), and Support Vector Machines (SVM). Normalising per stride produced significantly lower accuracies for SVM, LSTM and RF.

Table 4.23 Significant differences in mean accuracies for cross-validation test between normalisation method. Per = normalised per person

Model		Mean (%) ± *SD*		Mean (%) ± *SD*	Difference (%)	*p*-value
RF	Stride	78.69 ± 0.82	Raw	84.91 ± 0.74	6.22	<0.001
LSTM	Stride	76.53 ± 0.67	Per	80.99 ± 1.12	4.46	0.021
	Stride	76.53 ± 0.67	RP	81.00 ± 0.90	44.47	0.015
	Stride	76.53 ± 0.67	All	80.59 ± 0.75	4.06	0.048
SVM	Stride	61.93 ± 0.82	RP	71.36 ± 1.42	9.43	0.013
	Stride	61.93 ± 0.82	All	75.17 ± 1.09	14.24	<0.001
	Raw	65.62 ± 0.55	All	75.17 ± 1.09	9.55	0.015

Effects of types of prosthesis on terrain classification accuracies.

To see if separating the two types of prosthesis, transtibial (TT) and transfemoral (TF), will have an effect on the accuracy of the terrain classifier, trials were run with the data split by the prosthetic type. As this reduces the number of participants for the trials, the participants were split into 4 groups for the leave-some-out assessment. The groups were randomly assigned using the same method as done in section 4.2.13. There were twelve TT participants, three in each group, and nine TF participants, i.e. three groups of two and one group of three. The participants groups are displayed in Table 4.24.

Table 4.24 Participant groups for type of prosthesis trials

	Group 1	Group 2	Group 3	Group 4
TT	P1, P15, P18	P11, P13, P20	P6, P8, P17	P2, P7, P5
TF	P3, P16	P4, P12, P21	P10, P19	P9, P14

Table 4.25 and 4.26 display the mean accuracies and Figures 4.19 and 4.20 visualise the results for leave-some-out and cross-validation trials. There is no significant difference between the accuracies for leave-some-out trials from ANOVA tests (KNN $F(1, 2)$=2.03 p=0.18, RF $F(1, 2)$=0.22 p=0.81, SVM $F(1, 2)$=0.36 p=0.71, LSTM $F(1, 2)$=1.46 p=0.28). For cross-validation trials only SVM model showed statistical significance from Kruskal–Wallis tests (KNN $X^2(2)$=2.22 p=0.33, RF $X^2(2)$=3.86 p=0.15, SVM $X^2(2)$=10.82 p=0.005, LSTM $X^2(2)$=3.12 p=0.21). Dunn's post -hoc tests found that just classifying the TF data is significantly better than classifying all the data (TF M=77.76% SD= 1.71, All M=71.36% SD=1.42, difference=6.11% p=0.0034); however, these are the lowest cross-validation accuracies, and the other models showed no significant difference. Isolating TF does improve the mean accuracy for leave-some-out, but the variation is higher than when all participants are combined. It can therefore be said that separating data by prosthetic types does not improve the accuracies of the terrain classification which proves the null hypothesis.

Table 4.25 Leave-some-out terrain classification percentage accuracies for different prosthesis types. TT= Transtibial, TF = Transfemoral.

	KNN	RF	SVM	LSTM
TT	43.59	50.16	47.47	42.83
TF	50.46	53.29	52.49	44.68
Both	45.67%	53.81%	51.39%	48.77%

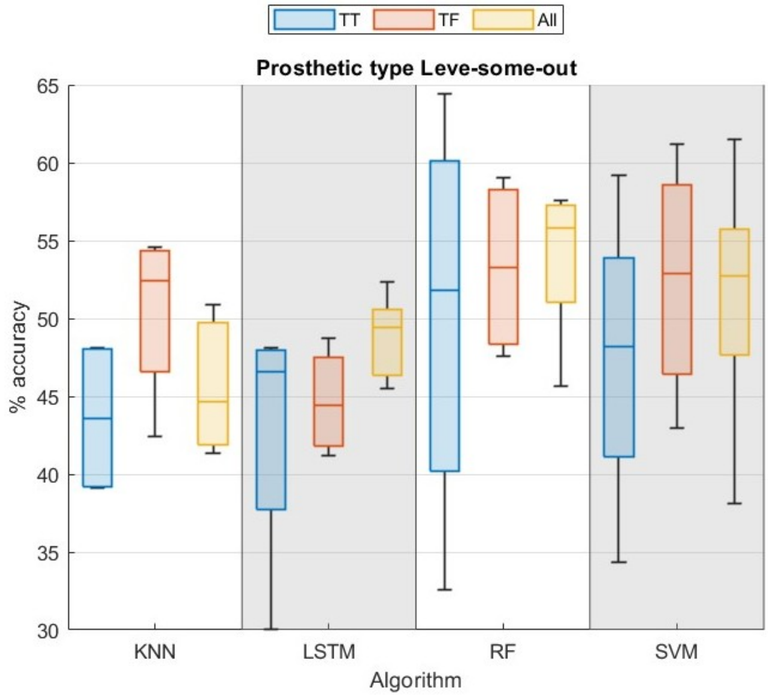

Fig. 4.19 Leave-some-out accuracy results for terrain classification when splitting different prosthetic types. The boxplot shows the classification accuracy for each prosthetic type (Transtibial-TT, Transfemoral-TF, both combined) using four machine learning algorithms: K-Nearest Neighbors (KNN), Long Short-Term Memory (LSTM), Random Forest (RF), and Support Vector Machines (SVM). There was no significant difference between models.

Table 4.26 Cross-validation terrain classification percentage accuracies for different prosthesis types. TT= Transtibial, TF = Transfemoral.

	KNN	RF	SVM	LSTM
TT	85.01%	82.91%	75.48%	79.80%
TF	86.45%	84.31%	77.47%	79.12%
Both	85.71%	83.00%	71.36%	81.00%

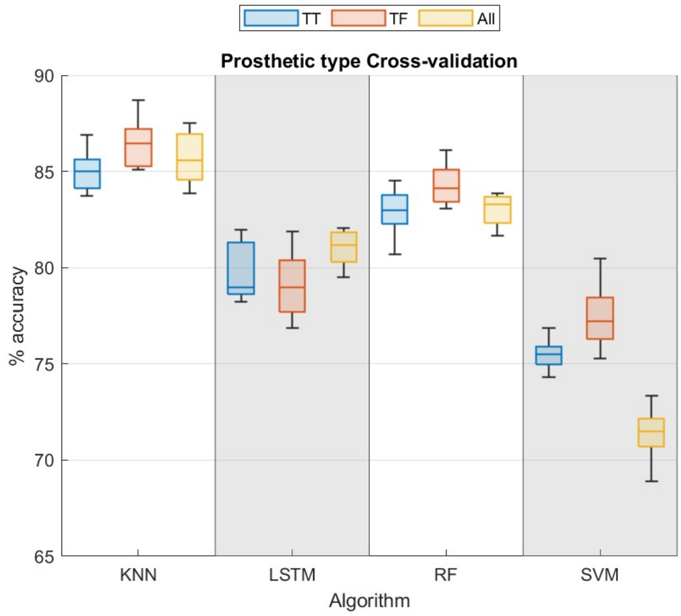

Fig. 4.20 Cross-validation accuracy results for terrain classification when splitting different prosthetic types. The boxplot shows the classification accuracy for each prosthetic type (Transtibial-TT, Transfemoral-TF, both combined) using four machine learning algorithms: K-Nearest Neighbors (KNN), Long Short-Term Memory (LSTM), Random Forest (RF), and Support Vector Machines (SVM). For SVM the combined produced lower accuracies but there was no significant difference for the other models.

Effect of walking aide use on terrain classification accuracies.

One of the final comparisons aimed to explore whether the classifiers work better if they know that a stride was taken with a walking aid or not. This was done by splitting the data into strides with a walking aid and without. Tables 4.27 and 4.28 display the mean accuracies and Figures 4.21 and 4.22 visualise the results. For leave-some-out, there was no significance between the results from Kruskal–Wallis tests (KNN $X^2(2)$=5.82 p=0.054, RF $X^2(2)$=2.06 p=0.36, SVM $X^2(2)$=1.46 p=0.48, LSTM $X^2(2)$=1.04 p=0.59). Isolating walking aid use did improve the mean accuracy but isolating strides without a walking aid decreased the accuracies. For cross-validation there was significance found in all models except RF from Kruskal–Wallis tests (KNN $X^2(2)$=6.86 p=0.032, RF $X^2(2)$=5.66 p=0.059, SVM $X^2(2)$=8.18 p=0.017, LSTM $X^2(2)$=12.02 p=0.003). Dunn's post-hoc tests found having the data combined produced better accuracies for all but SVM, with a significant improvement over classifying just the data with a walking aid for KNN (With M=83.24% SD=1.38, combined M=85.71% SD=1.31, difference=2.46% p=0.039) and LSTM (With M=73.19% SD=2.14, combined M=81.00% SD=0.90, difference=7.81% p=0.016), but isolating strides with a walking aid had significantly higher accuracies for SVM (With M=74.93% SD=1.53, combined M=71.36% SD=0.14, difference=3.57% p=0.022). Separating the data by walking aid use does not improve the accuracies of the classification models which proves the null hypothesis.

Table 4.27 Leave-one-out terrain classification percentage accuracies for trials with and without a walking aid

	KNN	RF	SVM	LSTM
With	49.18%	57.88%	53.98%	45.21%
Without	41.59%	51.01%	49.35%	42.68%
Both	45.67%	53.81%	51.39%	48.77%

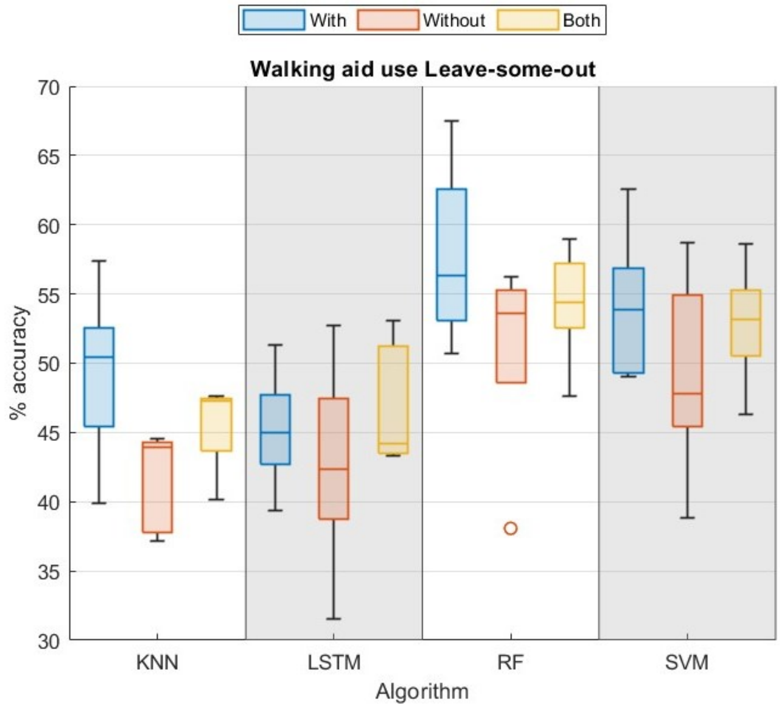

Fig. 4.21 Leave-some-out accuracy results for terrain classification across trials with and without a walking aid. The boxplot shows the classification accuracy for each variable (with, without and both together) using four machine learning algorithms: K-Nearest Neighbors (KNN), Long Short-Term Memory (LSTM), Random Forest (RF), and Support Vector Machines (SVM). There was no significant difference between the trials.

Table 4.28 Cross-validation terrain classification percentage accuracies for trials with and without a walking aid

	KNN	RF	SVM	LSTM
With	83.25%	81.16%	74.93%	73.19%
Without	83.38%	81.76%	71.54%	77.56%
Both	85.71%	83.00%	71.36%	81.00%

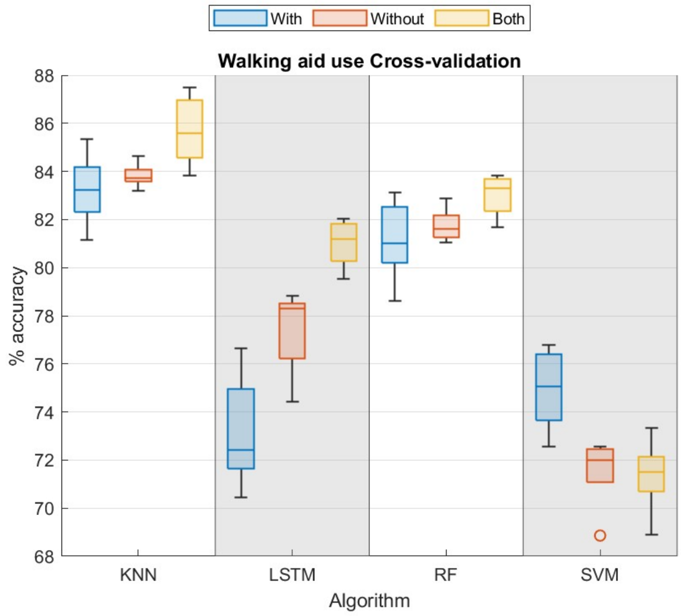

Fig. 4.22 Cross-validation accuracy results for terrain classification across trials with and without a walking aid. The boxplot shows the classification accuracy for each variable (with, without and both together) using four machine learning algorithms: K-Nearest Neighbors (KNN), Long Short-Term Memory (LSTM), Random Forest (RF), and Support Vector Machines (SVM). Data not separated produced higher accuracies for KNN, LSTM and RF but lower for SVM.

Effects of number of participants on terrain classification.

One of the limitations for this study is the limited number of participants (n=21) that were tested. Although it was higher than similar studies concerned with lower limb prosthetics [121][40][122], twenty-one participants may not have fully captured the variability in gait in the lower limb prosthetic user population. To investigate how the number of participants that are used to train the classifier affects the accuracy of the classifier at classifying data from unseen participants, a trial was run where the number of participants used to train the algorithms was changed from 1 to 20. The participants were randomly assigned to groups, and twenty-one groups were created for each trial. The mean accuracies for the trials are displayed in Table 4.29. The results are visualised in Figure 4.23. As hypothesised, increasing the number of participants increases the accuracies produced by the classifier. With the trial where twenty participants and one participant were used to train the algorithm, there was a larger variation in results compared to the other trials, as shown in Table 4.30, which demonstrates the variation in gait measures between participants. This larger variation in the trial where each participant was tested individually reduced the average accuracies, as seen as a slight plateau in the RF and LSTM plots in Figure 4.23. This indicated that with a larger participant pool higher accuracies could be achieved. The cross-validation results demonstrate the accuracies that could potentially be achieved by the terrain classifier if all the variations in gait could be captured in the training dataset.

Table 4.29 Terrain classification percentage accuracies for different number of participants used to train the classification models.

Participants used to train	KNN	RF	SVM	LSTM
20	42.90%	51.17%	50.73%	49.85%
18	40.51%	50.63%	48.69%	49.63%
14	37.66%	48.33%	46.85%	43.77%
11	36.83%	47.11%	45.81%	41.43%
7	34.69%	43.61%	43.25%	38.44%
3	31.42%	36.93%	36.25%	30.89%
1	26.19%	28.33%	28.05%	23.36%

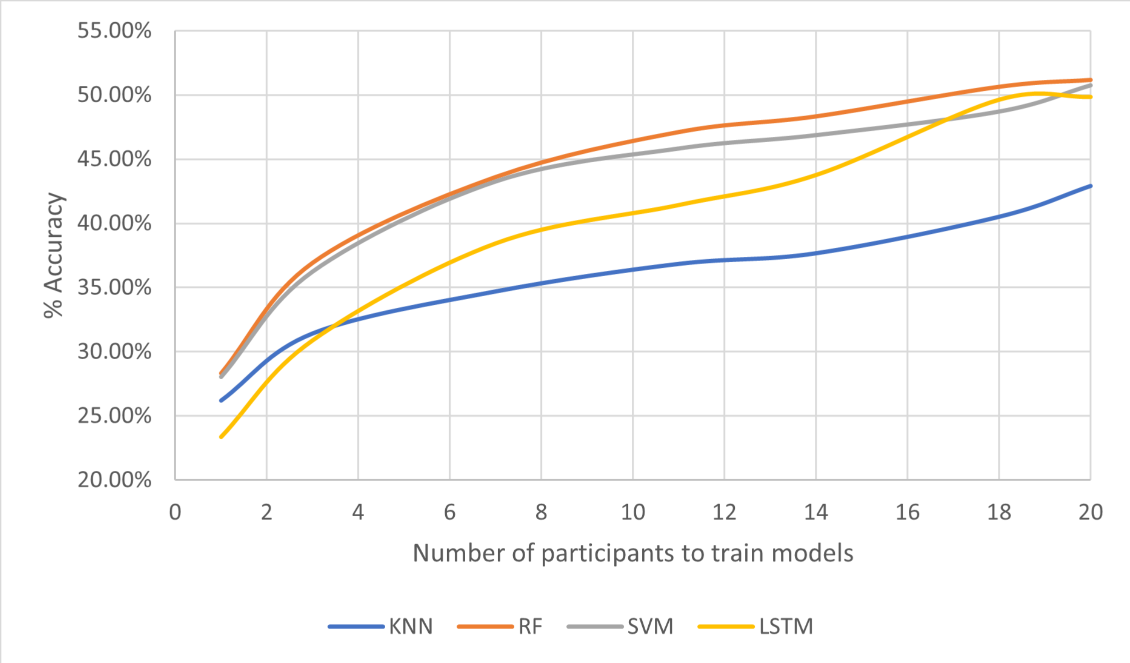

Fig. 4.23 Average percentage accuracy against the number of participants used to train the classification models. One to twenty participants were used to train four machine learning models: K-Nearest Neighbors (KNN), Long Short-Term Memory (LSTM), Random Forest (RF), and Support Vector Machines (SVM). Using more participants to train the models produced higher average percentage accuracies for all models.

Table 4.30 Variance of terrain classification percentage accuracies for different number of participants used to train the classification models.

Participants used to train	KNN	RF	SVM	LSTM
20	165.83	190.66	124.99	149.14
18	10.06	23.38	12.11	28.98
14	35.24	16.67	11.58	18.62
11	8.87	8.87	17.73	0.47
7	19.91	24.91	8.04	1.27
3	14.59	9.87	8.64	8.48
1	48.34	72.02	61.67	77.09

Final system informed by previous findings for terrain classification.

The following summarises the findings of the terrain classification trials:

- Using only a prosthetic shank-mounted IMU maintained high classification accuracy, demonstrating the system's efficiency without the need for additional sensors on the body.
- Splitting data into windows that cover an entire stride consistently provided the best classification performance, enhancing the system's robustness.
- Using 10 datapoints per window would reduce computational resource and data storage needs without affecting the classification accuracies.
- Velocity is a feature that can be removed to reduce the computational need whereas accelerations, gyroscope, magnetometer, and free accelerations produce high classification accuracies.
- Normalising per person, and combining this with the raw data, produces the best accuracies, whereas normalising by stride was shown to not improve classification accuracies, which, if eliminated, will reduce data processing.
- It was shown that splitting the data by prosthetic type or walking aid use does not improve classification accuracies, which reduces data processing.
- Using more participants to train the classifiers will improve the accuracies produced by the classifier.
- SVM and RF performed better with leave-some-out tests, whereas RF and KNN performed the best for cross-validation tests.

The chosen classifiers were run using 100 datapoints per stride to compare to the 10 datapoints per stride used for most of the analysis. Tables 4.31 and 4.32 display the 100 data point trials and 10 data point trials for leave-one-out and cross-validation and Figures 4.24 and 4.26 visualise the results. The 10 datapoint trials were statistically similar for all leave-one-out trials, but for cross-validation the 100 data point trial was significantly better for SVM and RF whilst significantly worse for KNN. The difference between the cross-validation RF is statistically significant but is only a 1.58% increase in mean accuracy. Using 10 datapoints would allow for collecting data at a lower sampling rate.

The 10 datapoint data were re-optimised, using the same methods as described in section 4.2.14, to see if this would improve the classification accuracies. For RF the hyperparameters that the optimisers output were: number of trees – 300, number of predictors – 70, minimum leaf size - 1. For KNN, the hyperparameters that the optimisers output were: number of neighbours – 1, how distance is calculated - Standardized Euclidean. For SVM, the hyperparameters that the optimisers output were: box constraint – 26.998, kernel size – 4.823, coding method – one vs all. The results for the optimised trials are shown in Tables 4.31 and 4.32 for the leave-some-out and cross-validation analysis, and comparison between the new optimised outcomes and the previous 10 datapoint results are visualised in Figures 4.25 and 4.27. The optimised trials were better for all cross-validation tests, but not significantly. For leave-some-out, the optimised trial was significantly worse for SVM running Dunn's test (Optimised M=30.20% SD=3.93, current M=51.39% SD=7.56, difference=21.19% p=0.0012). The optimiser works by looking at all the data whereas the leave-some-out trials exclude some participants from the training dataset to use as the test dataset. The reduced box constraint and kernel scale could increase the likelihood of overfitting which would affect the accuracies. The optimised RF leave-some-out trials produced the highest leave-some-out mean accuracy of all trials in this analysis. Due to this and the high cross-validation accuracy, the optimised RF classifier was used to classify the terrain in Chapter 6.

Table 4.31 Leave-some-out terrain classification percentage accuracies for trial with 100 and 10 datapoints and 10 datapoints with the models optimised.

	KNN	RF	SVM
100 datapoints	45.41%	54.62%	50.46%
10 datapoints	45.67%	53.81%	51.39%
10 datapoints optimised	46.61%	56.89%	30.20%

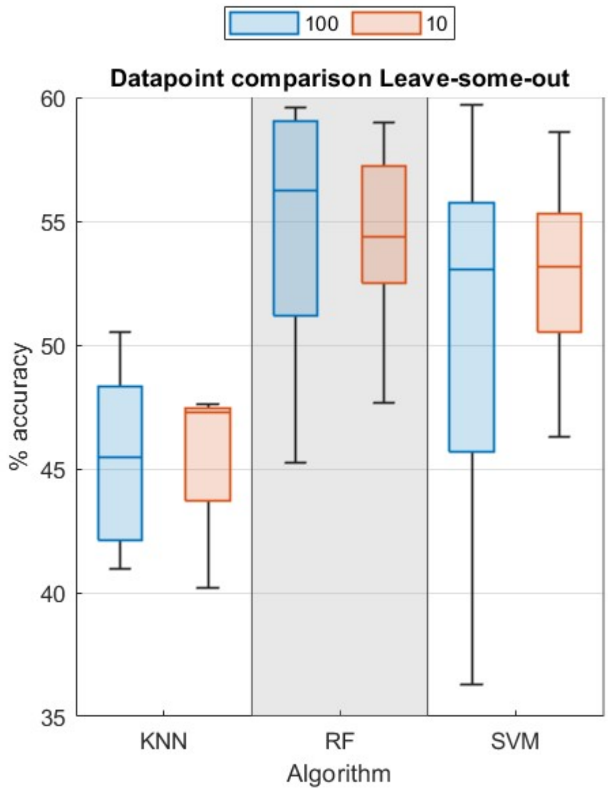

Fig. 4.24 Leave-some-out accuracy results for terrain classification across trials 10 and 100 datapoints. The boxplot shows the classification accuracy for 10 and 100 datapoints using the three best performing machine learning algorithms from the previous analysis: K-Nearest Neighbors (KNN), Random Forest (RF), and Support Vector Machines (SVM). Using 10 datapoints was not significantly different to 100 datapoints.

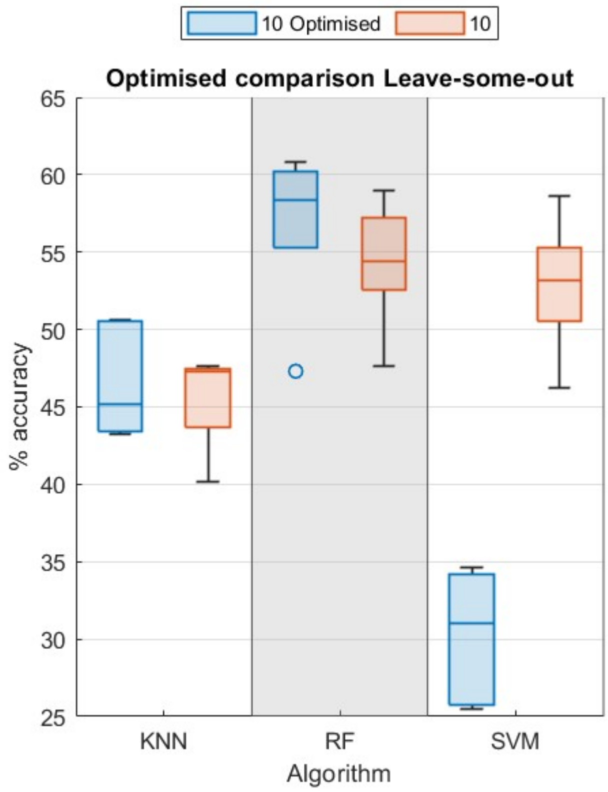

Fig. 4.25 Leave-some-out accuracy results for terrain classification across trials with 10 datapoints using models with the hyperparameters that have been used for all of the analysis and models that have had their hyperparameters optimised. The boxplot shows the classification accuracy for the two types of models (10 -using hyperparameters used in previous analysis, 10 optimised – using optimised hyperparameters) using the three best performing machine learning algorithms from the previous analysis: K-Nearest Neighbors (KNN), Random Forest (RF), and Support Vector Machines (SVM). The optimised trials produced a significantly lower accuracy for SVM, but this was the model that produced lower accuracies.

Table 4.32 Cross-validation terrain classification percentage accuracies for trial with 100 and 10 datapoints and 10 datapoints with the models optimised.

	KNN	RF	SVM
100 datapoints	82.20%	84.58%	84.97%
10 datapoints	85.71%	83.00%	71.36%
10 datapoints optimised	85.65%	83.64%	73.00%

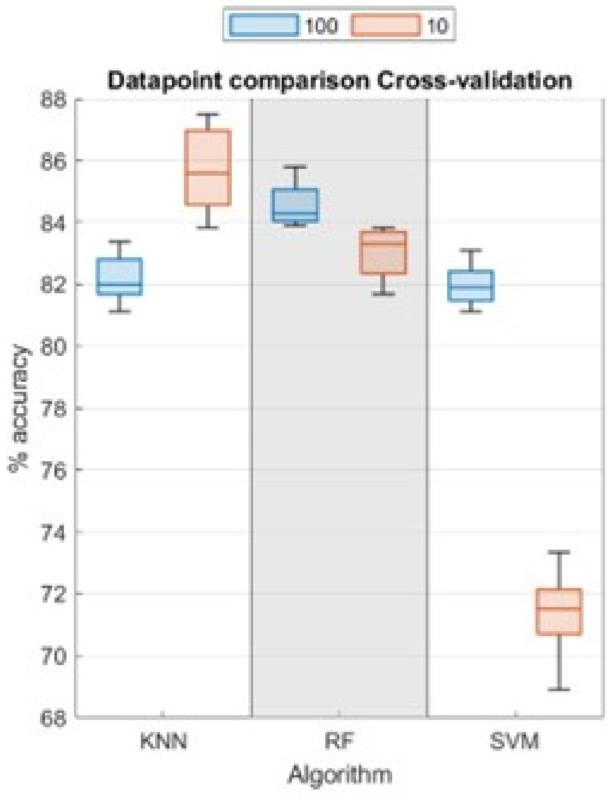

Fig. 4.26 Cross-validation accuracy results for terrain classification across trials 10 and 100 datapoints. The boxplot shows the classification accuracy for 10 and 100 datapoints using the three best performing machine learning algorithms from the previous analysis: K-Nearest Neighbors (KNN), Random Forest (RF), and Support Vector Machines (SVM). Using 10 datapoints only produced significantly lower accuracies with the SVM model.

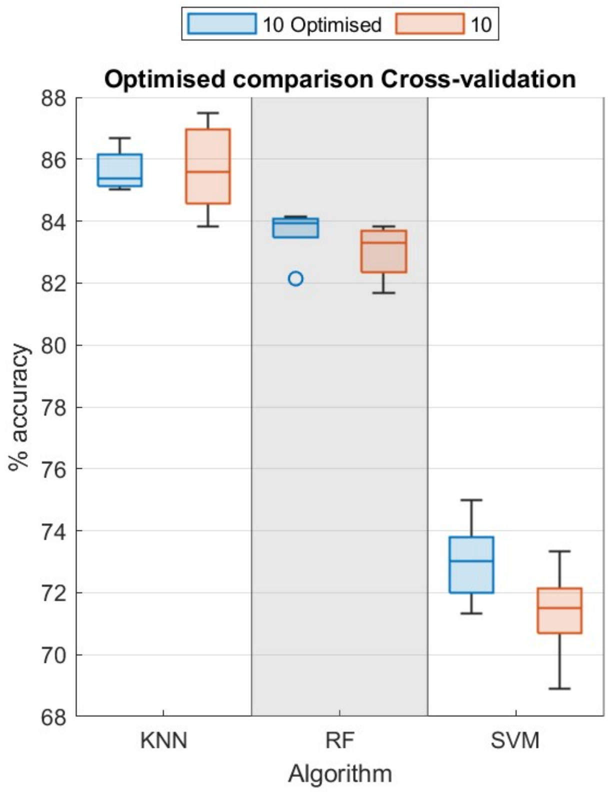

Fig. 4.27 Cross-validation accuracy results for terrain classification across trials with 10 datapoints using models with the hyperparameters that have been used for all of the analysis and models that have had their hyperparameters optimised. The boxplot shows the classification accuracy for the two types of models (10 -using hyperparameters used in previous analysis, 10 optimised – using optimised hyperparameters) using the three best performing machine learning algorithms from the previous analysis: K-Nearest Neighbors (KNN), Random Forest (RF), and Support Vector Machines (SVM). The optimised trials produced higher mean accuracies but not to a significant level.

Following this detailed investigation of the effects on terrain classification accuracies in relation to use of sensors, windows, data points per window, variables, normalization, prosthesis type and terrain/walking aid use, effects of the same on classification of walking aid use will be discussed in the subsequent sections.

4.3.3 Effects on 'walking aid use' classification accuracies.

This section investigates walking aid recognition from the IMU data using machine learning algorithms. The same style of analysis was run on the walking aid use recognition models as for the terrain classification. The only difference was that LR models were also assessed because LR models look at binary results such as walking aid use/non-use and give a baseline accuracy that other models can be compared against.

Effects of sensor location and combination on walking aid use recognition accuracies.

As before, to investigate how IMUs in the different locations can recognise walking aid use, trials were run for each IMU with 100Hz data split into strides of 100 datapoints. The mean leave-some-out and cross-validation accuracies are displayed in Tables 4.33 and 4.34, and the results visualised in Figures 4.28 and 4.29. Trunk and the combined trials produced the higher accuracies for KNN, RF and SVM for both leave-some-out and cross-validation, but for LR and LSTM produced lower accuracies. There were no significant differences between the leave-some-out trials from Kruskal–Wallis tests (KNN $X^2(4)=5.29$ $p=0.26$, RF $X^2(4)=0.28$ $p=0.99$, SVM $X^2(4)=1.41$ $p=0.84$, LR $X^2(4)=4.86$ $p=0.30$, LSTM $X^2(4)=3.27$ $p=0.51$). There were significant differences, however, for all the cross-validation trials from Kruskal–Wallis tests (KNN $X^2(4)=18.35$ $p=0.0011$, RF $X^2(4)=12.87$ $p=0.012$, SVM $X^2(4)=20.53$ $p<0.0001$, LR $X^2(4)=15.49$ $p=0.0038$, LSTM $X^2(4)=19.14$ $p<0.0001$). Results for Dunn's post hoc significance testing are displayed in Table 4.35 The prosthetic shank IMU was significantly worse than the trunk and the combined IMUs for KNN. For RF, the other shank IMU was significantly worse than the combined IMUs. For SVM, the other shank and trunk IMUs were significantly better than combined, and the prosthetic shank IMU was significantly worse than the trunk IMU. For the LR trials, the combined IMUs performed significantly worse than the prosthetic shank, other shank and thigh IMUs. The combined IMUs also performed significantly worse for the LSTM trials to the other shank and thigh IMUs. These results disprove the null hypothesis that IMU location does not have an effect on walking aid recognition accuracies. As the ideal system would be solely prosthetic mounted, there is no significant difference using the prosthetic shank IMU to any other IMU location when looking at leave-some-out data, and although the prosthetic shank IMU performs statistically significantly worse for KNN and SVM trials, the actual mean accuracies are only slightly smaller (KNN 2.5±1%, SVM 1.9±0.6%), hence only the prosthetic shank IMU will be used for the rest of the analysis.

Table 4.33 Leave-some-out walking aid recognition percentage accuracies for different IMU positions. PS = Prosthetic shank, OS = Other shank, TR = Trunk, TH = Thigh.

	KNN	RF	SVM	LR	LSTM
PS	55.08%	61.45%	64.39%	53.70%	54.59%
OS	52.44%	61.43%	62.58%	55.35%	56.55%
TR	54.83%	65.5%	67.95%	55.65%	59.84%
TH	53.48%	60.68%	60.61%	48.43%	56.49%
All	51.40%	66.50%	67.45	48.99%	64.45%

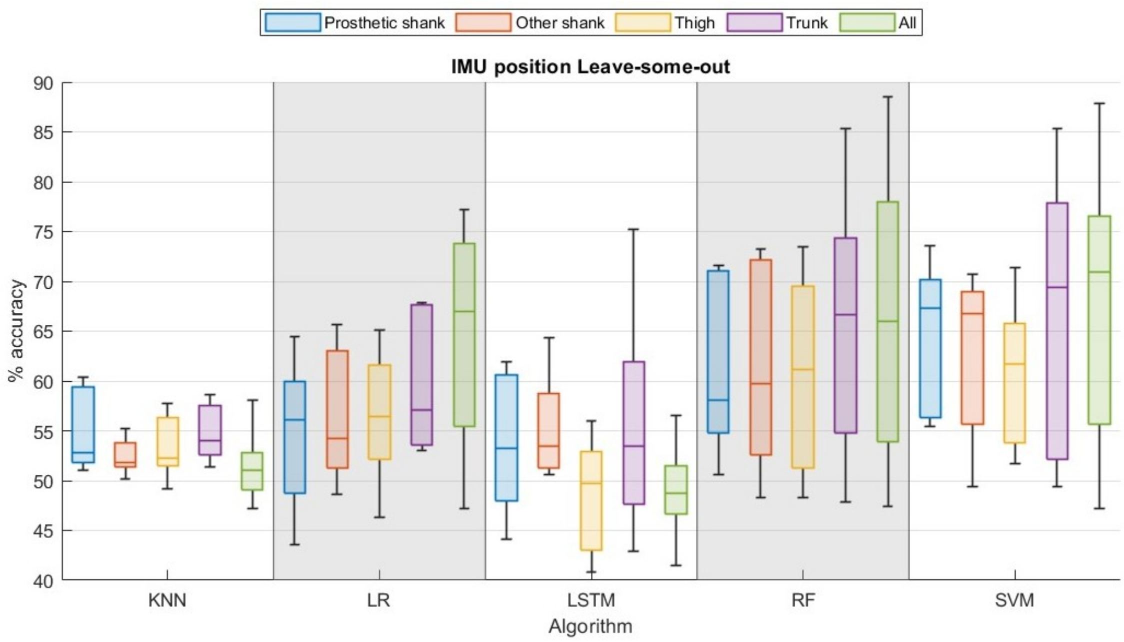

Fig. 4.28 Leave-some-out accuracy results for walking aid recognition for different IMU positions. The boxplot shows the classification accuracy for each IMU position (PS = Prosthetic shank, OS = Other shank, TR = Trunk, TH = Thigh, all combined) using four machine learning algorithms: K-Nearest Neighbors (KNN), Long Short-Term Memory (LSTM), Random Forest (RF), and Support Vector Machines (SVM). There was no significant difference for all models.

Table 4.34 Cross-validation walking aid recognition percentage accuracies for different IMU positions. PS = Prosthetic shank, OS = Other shank, TR = Trunk, TH = Thigh.

	KNN	RF	SVM	LR	LSTM
PS	74.31%	79.39%	82.21%	65.15%	69.29%
OS	75.22%	79.11%	83.97%	65.76%	70.79%
TR	77.24%	80.29%	84.84%	58.01%	68.91%
TH	76.63%	79.41%	83.45%	65.24%	71.43%
all	78.21%	81.71%	75.04%	53.19%	65.61%

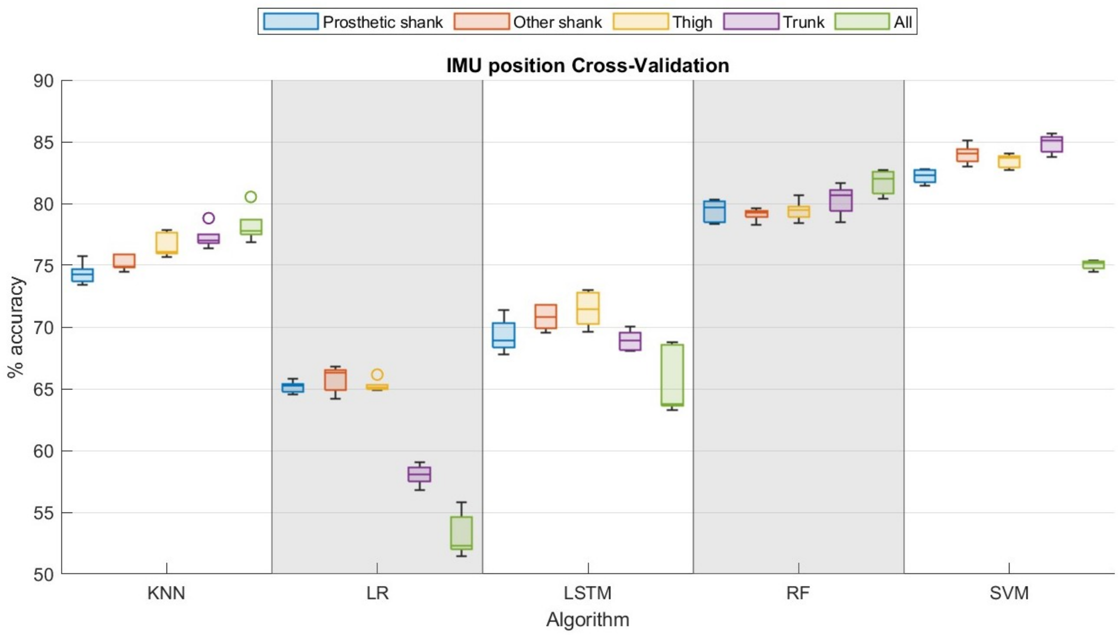

Fig. 4.29 Cross-validation accuracy results for walking aid recognition for different IMU positions. The boxplot shows the classification accuracy for each IMU position (PS = Prosthetic shank, OS = Other shank, TR = Trunk, TH = Thigh, all combined) using four machine learning algorithms: K-Nearest Neighbors (KNN), Long Short-Term Memory (LSTM), Random Forest (RF), and Support Vector Machines (SVM). PS produces lower accuracies to the TR and combined IMUs for KNN, and Trunk for SVM. OS produced lower accuracies than the combined IMUs for RF. The combined IMUs produced lower accuracies than PS, OS and TH IMUs for LR, OT and TR for SVM, and OS and TH for LSTM.

Table 4.35 Significant differences in mean accuracies for walking aid recognition cross-validation test between IMU locations

Model		Mean (%) ± *SD*		Mean (%) ± *SD*	Difference (%)	*p*-value
KNN	PS	74.31 ± 0.80	Trunk	77.24 ± 0.83	2.93	0.026
	PS	74.31 ± 0.80	All	78.21 ± 1.23	3.90	0.0026
RF	OS	79.11 ± 0.45	All	81.71 ± 0.90	2.60	0.017
SVM	All	75.04 ± 0.34	OS	83.97 ± 0.69	8.93	0.02
	All	75.04 ± 0.34	Trunk	84.84 ± 0.68	9.80	<0.001
	PS	82.21 ± 0.49	Trunk	84.84 ± 0.68	2.63	0.034
LR	All	53.19 ± 1.60	PS	65.15 ± 0.43	11.96	0.026
	All	53.19 ± 1.60	OS	65.76 ± 0.96	12.57	0.002
	All	53.19 ± 1.60	Thigh	65.24 ± 0.46	12.05	0.03
LSTM	All	65.61 ± 2.48	OS	70.79 ± 0.92	5.18	0.026
	All	65.61 ± 2.48	Thigh	71.43 ± 1.29	5.82	0.007

Effect of window type on walking aid use recognition accuracies.

Trials were run to see how different window types affect walking aid recognition accuracies. Tables 4.36 and 4.37 display the mean accuracies for each window type using each classifier, and Figures 4.30 and 4.31 visualise the results. The leave-some-out trials produced statistically similar accuracies for all the window types for each classifier from Kruskal–Wallis tests (KNN $X^2(7)$=7.02 p=0.43, RF $X^2(7)$=1.24 p=0.99, SVM $X^2(7)$=5.12 p=0.64, LR $X^2(7)$=5.69 p=0.58, LSTM $X^2(7)$=3.22 p=0.86). For the cross-validation tests all models produced statistically significant results from Kruskal–Wallis tests (KNN $X^2(6)$=32.01 p<0.001, RF $X^2(6)$=33.16 p<0.001, SVM $X^2(6)$=32.98 p<0.001, LR $X^2(6)$=31.82 p<0.001, LSTM $X^2(7)$=25.13 p<0.001). From Dunn's post-hoc tests, the stride-window produced significantly higher accuracies than the 2s, and 1.5s time-windows for RF, 2s, 1.5s and 0.5s time-windows for SVM and the 2s, 1.5s and 0.2s time-windows for LSTM. For RF, the 0.1s time-window produced significantly higher accuracies then the 2s, 1.5s and 0.5s time-windows. The 0.1s time-window also produced significantly higher accuracies then the 2s and 0.5s time windows for KNN, and the 2s time window for SVM and LR. The 1s time window produced significantly higher accuracies than the 2s and 0.5s for KNN, the 2s and 1.5s for LR, and 2s for SVM. Results for significance testing are displayed in Table 4.38. These results disprove the null hypothesis that windowing method does not have an effect on walking aid recognition accuracies. The highest accuracies for both the leave-some-out and cross-validation being produced by splitting the data into strides in the SVM test, this was the method used for the rest of the analysis.

Table 4.36 Leave-some-out walking aid recognition percentage accuracies for different window types.

	KNN	RF	SVM	LR	LSTM
0.1s	58.12%	61.64%	58.64%	63.46%	61.55%
0.2s	57.88%	61.81%	59.01%	54.85%	61.68%
0.5s	56.08%	62.19%	60.63%	48.13%	59.14%
1s	53.89%	61.55%	62.68%	50.91%	60.79%
1.5s	52.77%	61.15%	61.96%	49.93%	59.52%
2s	51.99%	61.42%	61.46%	50.73%	57.48%
Step	55.08%	61.45%	64.39%	53.70%	54.59%
Individual	59.52%	59.24%	62.65%	54.39%	60.43%

Table 4.37 Cross-validation walking aid recognition percentage accuracies for different window types.

	KNN	RF	SVM	LR	LSTM
0.1s	77.96%	81.88%	79.53%	65.64%	62.18%
0.2s	73.55%	76.96%	77.19%	55.69%	61.05%
0.5s	69.20%	72.90%	72.01%	56.23%	62.74%
1s	77.78%	74.96%	78.61%	72.35%	66.24%
1.5s	70.56%	66.06%	68.97%	50.65%	60.20%
2s	64.40%	68.81%	65.87%	51.22%	59.12%
Step	74.31%	79.39%	82.22%	65.15%	69.29%

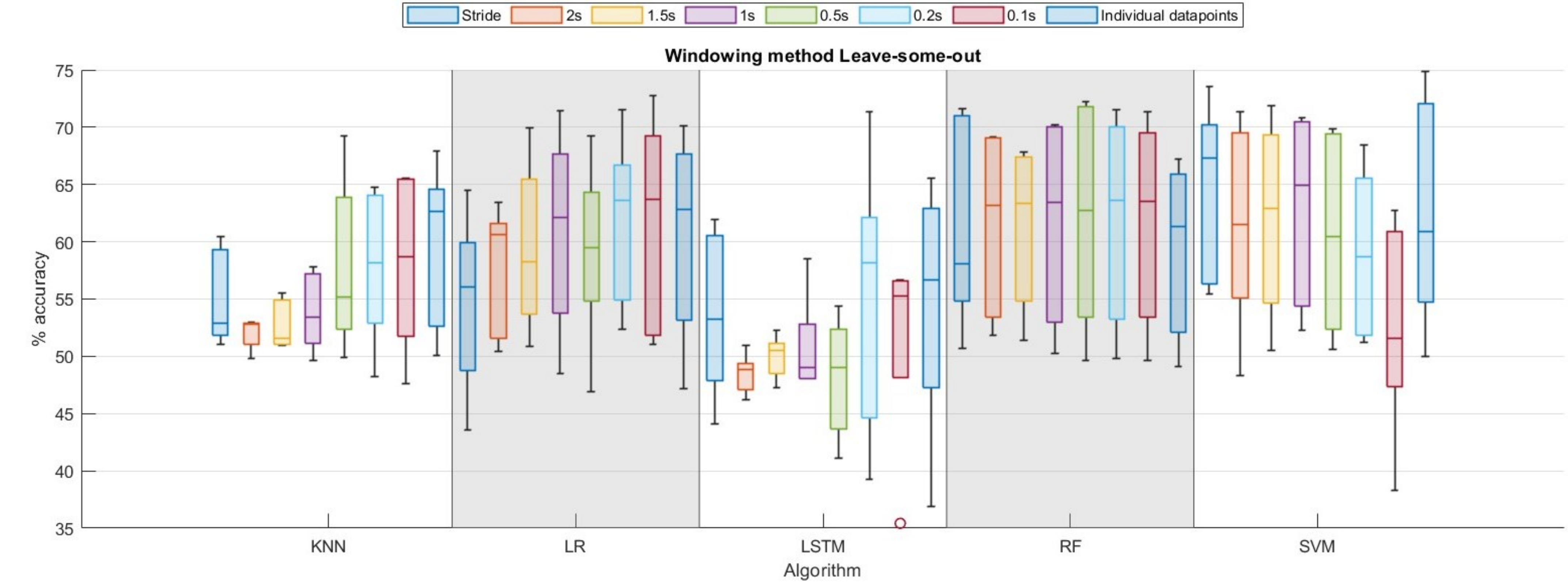

Fig. 4.30 Leave-some-out accuracy results for walking aid recognition across different windowing methods. The boxplot shows the classification accuracy for windowing method (time-based windows 0.1s, 0.2s, 0.5s, 1s, 1.5s and 2s, stride-based window and individual datapoints) using four machine learning algorithms: K-Nearest Neighbors (KNN), Long Short-Term Memory (LSTM), Random Forest (RF), and Support Vector Machines (SVM). There was no significant different between methods.

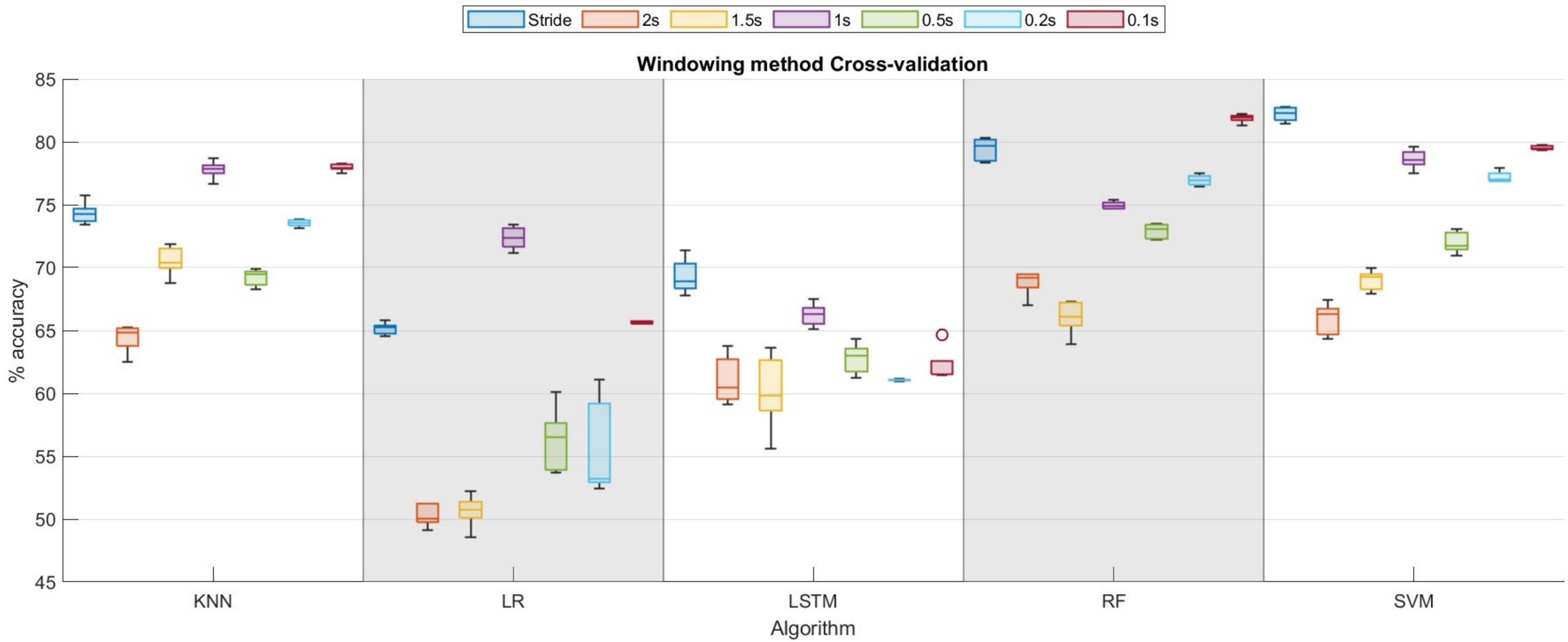

Fig. 4.31 Cross-validation accuracy results for walking aid recognition across different windowing methods. The boxplot shows the classification accuracy for windowing method (time-based windows 0.1s, 0.2s, 0.5s, 1s, 1.5s and 2s, stride-based window and individual datapoints) using four machine learning algorithms: K-Nearest Neighbors (KNN), Long Short-Term Memory (LSTM), Random Forest (RF), and Support Vector Machines (SVM). The stride-based window produced the highest accuracy in the SVM test.

Table 4.38 Significant differences in mean accuracies for walking aid recognition cross-validation test between windowing methods

Model		Mean (%) ± *SD*		Mean (%) ± *SD*	Difference (%)	*p*-value
KNN	0.1s	77.96 ± 0.26	2s	64.40 ± 1.01	13.56	<0.001
	0.1s	77.96 ± 0.26	0.5s	69.20 ± 0.59	8.76	0.013
	1s	77.78 ± 0.66	2s	64.40 ± 1.01	13.38	<0.001
	1s	77.78 ± 0.66	0.5s	69.20 ± 0.59	8.58	0.022
RF	Stride	79.39 ± 0.82	2s	68.81 ± 1.01	10.58	0.034
	Stride	79.39 ± 0.82	1.5s	66.06 ± 1.22	13.33	0.003
	0.1s	81.88 ± 0.31	2s	68.81 ± 1.01	13.07	0.002
	0.1s	81.88 ± 0.31	1.5s	66.06 ± 1.22	11.32	<0.001
	0.1s	81.88 ± 0.31	0.5s	72.90 ± 0.54	8.98	0.042
SVM	Stride	82.21 ± 0.49	2s	65.87 ± 1.15	16.34	<0.001
	Stride	82.21 ± 0.49	1.5s	68.97 ± 0.72	13.25	0.002
	Stride	82.21 ± 0.49	0.5s	72.01 ± 0.78	10.21	0.042
	0.1s	79.53 ± 0.16	2s	65.87 ± 1.15	13.66	0.003
	1s	78.61 ± 0.70	2s	65.87 ± 1.15	12.74	<0.001
LSTM	Stride	69.29 ± 1.26	2s	61.09 ± 1.73	10.17	0.013
	Stride	69.29 ± 1.26	1.5	60.20 ± 2.75	9.09	0.007
	Stride	69.29 ± 1.26	0.2s	61.05 ± 0.07	8.24	0.004
LR	0.1s	65.64 ± 0.09	2s	50.32 ± 0.82	14.42	0.016
	1s	72.35 ± 0.81	2s	50.32 ± 0.82	21.13	<0.001
	1s	72.35 ± 0.81	1.5s	50.65 ± 1.18	21.70	<0.001

Effects of number of datapoints per window on walking aid use recognition accuracies.

As splitting the data into windows that contain a whole stride performed better than the other window types, this is the window type that will be used for the rest of the analysis. As with the terrain classification, to understand if the number of datapoints affects the accuracies of the classifiers, trials were run with windows of 10, 20, 50 and 100 datapoints. Table 4.39 and 4.40 display the mean accuracies and Figures 4.32 and 4.33 visualise the results. For leave-some-out trials none of the models produced a statistically significant result from Kruskal–Wallis tests (KNN $X^2(3)$=0.55 p=0.91, RF $X^2(3)$=0.55 p=0.91, SVM $X^2(3)$=1.86 p=0.60, LR $X^2(3)$=0.19 p=0.98, LSTM $X^2(3)$=0.70 p=0.87). For cross-validation trials only LR produced a statistically significant result from Kruskal–Wallis tests (KNN $X^2(3)$=1.84 p=0.61, RF $X^2(3)$=2.29 p=0.51, SVM $X^2(3)$=6.19 p=0.10, LR $X^2(3)$=15.48 p=0.001, LSTM $X^2(3)$=3.62 p=0.31). Dunn's post-hoc tests showed the only significant difference for the cross-validation LR accuracy when splitting the data into 100 datapoints which performed significantly worse than 20 datapoint splits (100 datapoint M=65.15% SD=0.43, 20 datapoints M=73.25% SD=0.55, difference=8.10% p<0.001). These results prove the null hypothesis that the number of datapoints in a window does not effect walking aid recognition accuracies. As there is no significant reduction in accuracy using 10 datapoints in any of the trials, this will reduce the computational need to run the analysis, and hence strides split into 10 datapoints will be used for the remaining analysis.

Table 4.39 Leave-some-out walking aid recognition percentage accuracies for different numbers of datapoints per window.

	KNN	RF	SVM	LR	LSTM
10	55.97%	60.07%	59.61%	53.61%	57.02%
20	55.17%	62.07%	60.22%	54.64%	55.96%
50	54.69%	61.99%	62.74%	50.90%	55.57%
100	55.08%	61.45%	64.39%	53.7%	54.59%

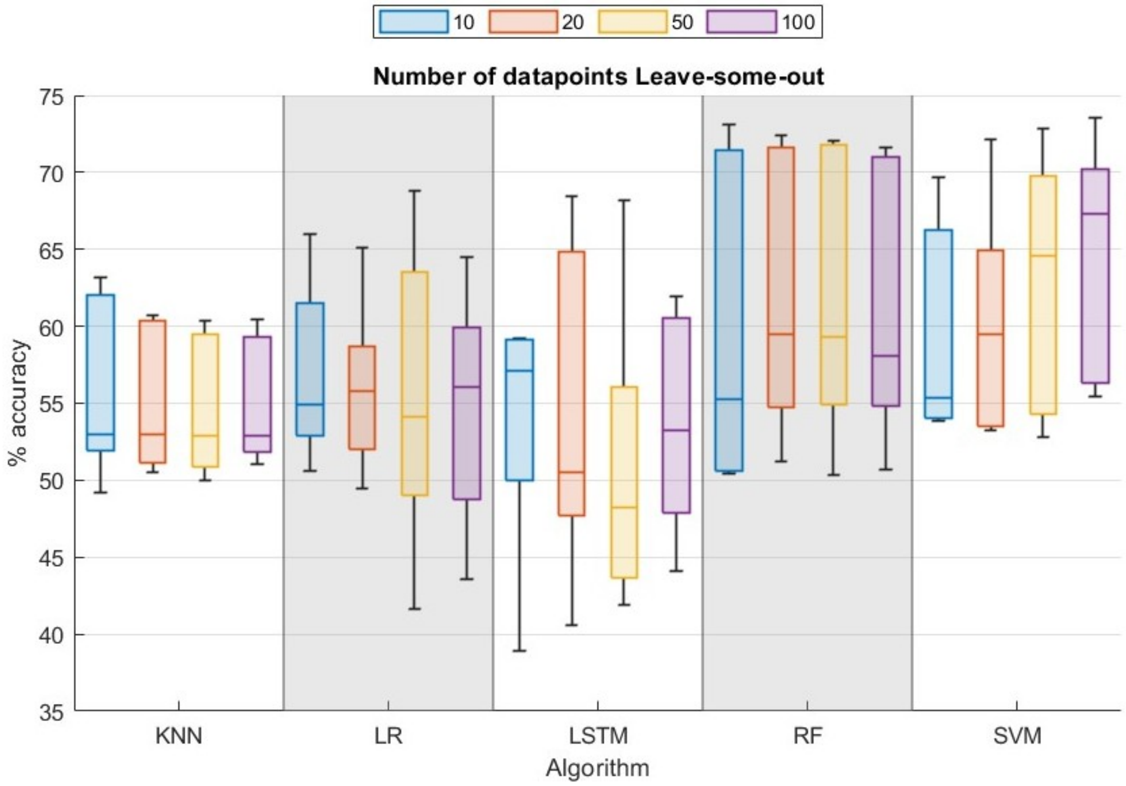

Fig. 4.32 Leave-some-out accuracy results for walking aid recognition for number of datapoints in a window. The boxplot shows the classification accuracy for each number of datapoints (10, 20, 50, 100) using four machine learning algorithms: K-Nearest Neighbors (KNN), Long Short-Term Memory (LSTM), Random Forest (RF), and Support Vector Machines (SVM). There was no significant difference for all models.

Table 4.40 Cross-validation walking aid recognition percentage accuracies for different numbers of datapoints per window.

	KNN	RF	SVM	LR	LSTM
10	74.85%	78.71%	83.10%	70.86%	66.66%
20	74.79%	79.60%	83.32%	73.25%	67.75%
50	74.60%	79.25%	83.28%	71.78%	67.39%
100	74.31%	79.39%	82.22%	65.15%	69.29%

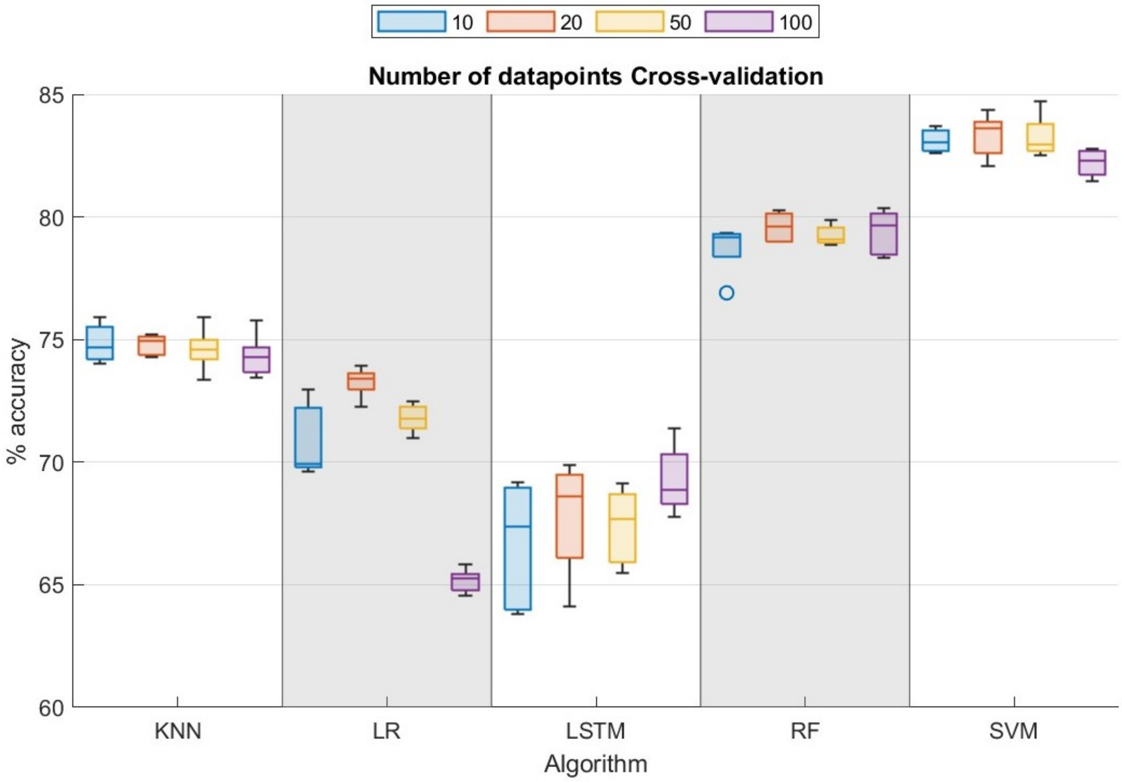

Fig. 4.33 Cross-validation accuracy results for walking aid recognition for number of datapoints in a window. The boxplot shows the classification accuracy for each number of datapoints (10, 20, 50, 100) using four machine learning algorithms: K-Nearest Neighbors (KNN), Long Short-Term Memory (LSTM), Random Forest (RF), and Support Vector Machines (SVM). The 100 datapoint window trial using LR produced lower accuracies than the 20 datapoint window. There was no significant difference between the other trials.

Effects of variables used on walking aid use recognition accuracies.

Another way to reduce the computational need to recognise walking aid use is to reduce the features (i.e. variables) that are used by the different models. Identifying irrelevant features could also simplify the sensor system by eliminating the need for certain measures to be recorded. As with the terrain classification, section 4.3.2, each feature was run separately to see how the accuracies were affected. The features that produced the highest mean accuracies were then combined and run in separate trials to see how combining certain variables affects the accuracies. There was no significant difference between the leave-some-out trials from Kruskal–Wallis tests (KNN $X^2(8)=2.59$ $p=0.96$, RF $X^2(8)=3.75$ $p=0.88$, SVM $X^2(8)=4.52$ $p=0.81$, LR $X^2(8)=4.26$ $p=0.83$, LSTM $X^2(8)=2.94$ $p=0.94$). For the cross-validation, tests showed significance for all the classifiers from Kruskal–Wallis tests (KNN $X^2(8)=35.51$ $p<0.001$, RF $X^2(8)=38.09$ $p<0.001$, SVM $X^2(8)=42.18$ $p<0.001$, LR $X^2(8)=18.50$ $p=0.018$, LSTM $X^2(8)=32.48$ $p<0.001$). Dunn's post-hoc tests showed KNN magnetometer and cadence were significantly worse than AGV and AGVC. For RF, magnetometer and cadence produced significantly lower accuracies to AGV and AGVC, cadence was also significantly lower than all the variable combined. For SVM, magnetometer and cadence produce significantly lower accuracies than AGV, AGVC and all the variables combined. Free accelerations also produced lower accuracies than AVGC. Cadence produced significantly lower accuracies than accelerations, AVG, AVGC and all the variables combined for LR. Results for significance testing are displayed in Table 4.43.. All the combined measures were statistically similar for all models. The combination of accelerations, gyroscope, velocities and cadence produced the highest mean accuracies for seven out of the ten algorithms, and this included SVM which had the highest mean accuracies for both verification techniques. AVGC was therefore used for the subsequent analysis. These results disprove the null hypothesis that the variables used to classify walking aid use have no effect on the walking aid recognition accuracies.

Table 4.41 Leave-some-out walking aid recognition percentage accuracies for different variables. AGV = accelerations, gyroscope and velocity, AGVC = accelerations, gyroscope, velocity and cadence.

	KNN	RF	SVM	LR	LSTM
Accelerations	54.81%	65.04%	59.76%	59.58%	59.56%
Gyroscope	57.40%	55.68%	58.39%	56.98%	55.47%
Magnetometer	54.65%	60.66%	55.32%	57.63%	55.03%
Free Accelerations	55.25%	61.33%	59.24%	56.47%	57.02%
Cadence	55.83%	59.59%	61.53%	62.84%	61.82%
Velocity	54.94%	61.04%	59.43%	54.88%	53.75%
AGV	57.42%	61.75%	62.67%	56.12%	55.13%
AGVC	57.93%	61.36%	63.83%	57.80%	61.55%
All	55.97%	60.07%	59.61%	53.61%	57.02%

Table 4.42 Cross-validation walking aid recognition percentage accuracies for different variables. AGV = accelerations, gyroscope and velocity, AGVC = accelerations, gyroscope, velocity and cadence.

	KNN	RF	SVM	LR	LSTM
Accelerations	74.96%	77.84%	78.94%	72.36%	66.94%
Gyroscope	75.48%	76.80%	76.54%	68.23%	63.34%
Magnetometer	70.00%	72.90%	68.59%	67.02%	62.32%
Free Accelerations	73.56%	73.92%	73.87%	66.99%	64.27%
Cadence	65.66%	66.39%	63.08%	63.05%	62.94%
Velocity	75.51%	77.88%	76.79%	69.06%	65.73%
AGV	76.35%	79.06%	83.51%	73.23%	63.78%
AGVC	76.67%	79.21%	84.16%	71.72%	65.65%
All	74.85%	78.71%	83.10%	70.86%	66.66%

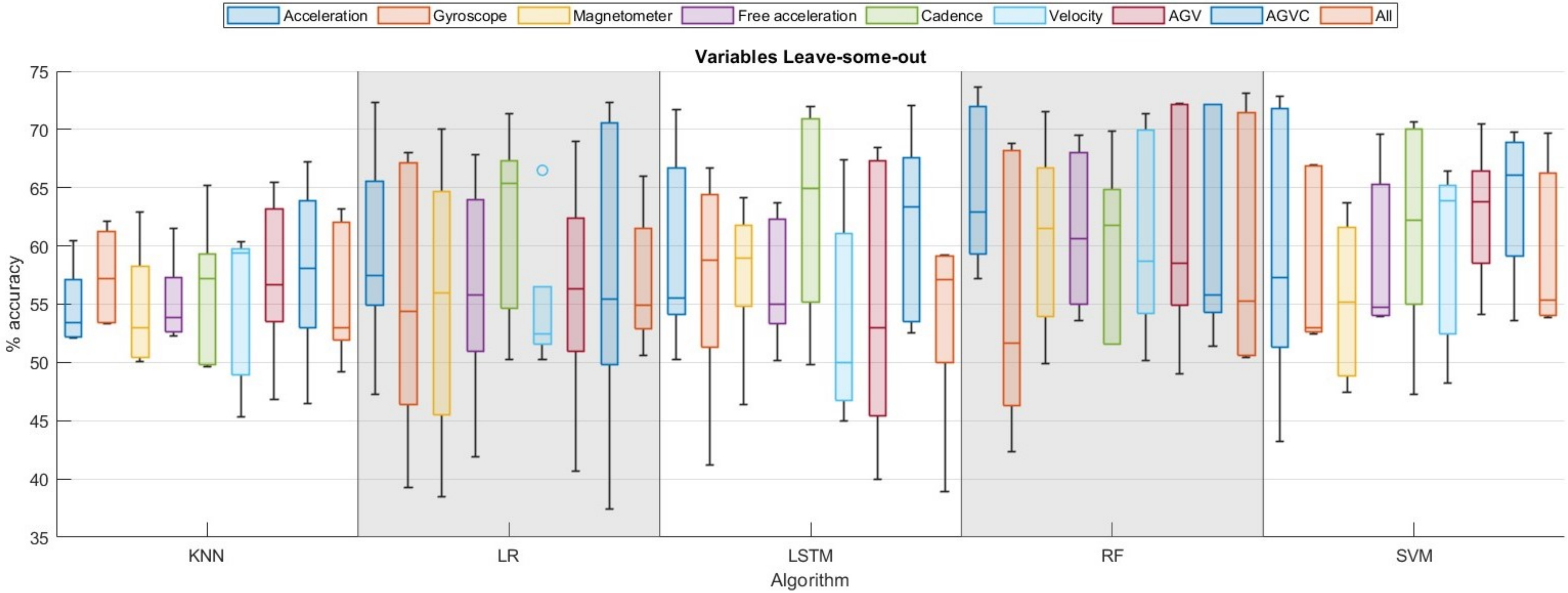

Fig. 4.34 Leave-some-out accuracy results for walking aid recognition across different variables and variable groups. The boxplot shows the classification accuracy for each variable (acceleration, gyroscope, magnetometer, free accelerations, cadence, velocity, AGV = accelerations, gyroscope and velocity, AGVC = accelerations, gyroscope, velocity and cadence, and all variables combined) using four machine learning algorithms: K-Nearest Neighbors (KNN), Long Short-Term Memory (LSTM), Random Forest (RF), and Support Vector Machines (SVM). There was no significant difference between the results.

data.

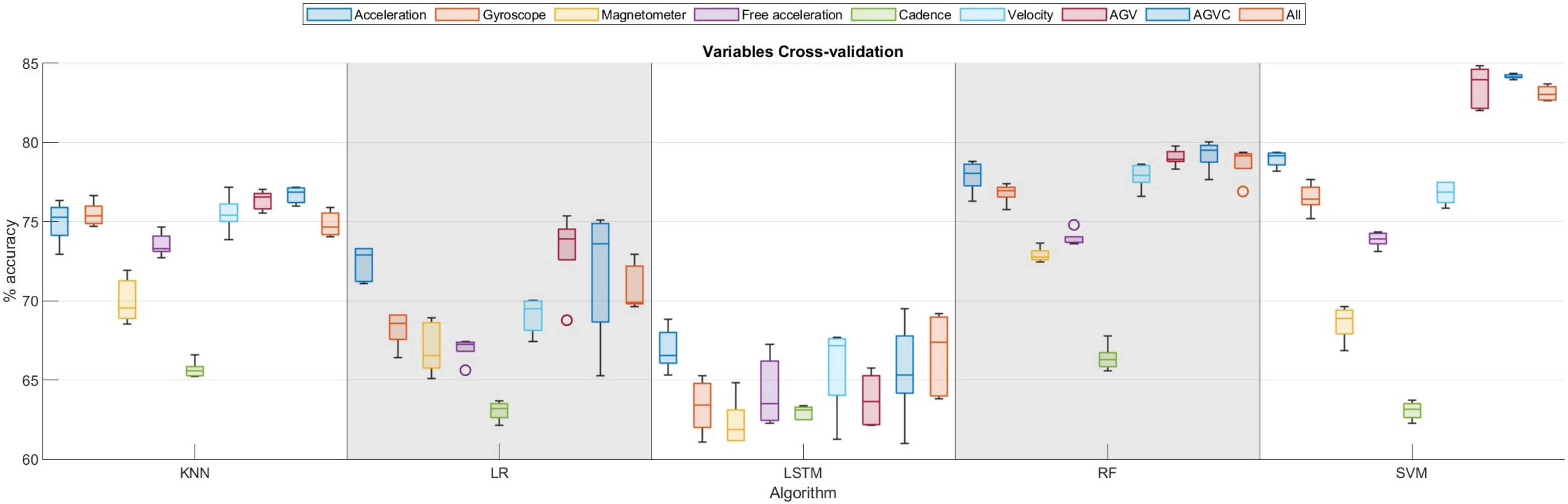

Fig. 4.35 Cross-validation accuracy results for walking aid recognition across different variables and variable groups. The boxplot shows the classification accuracy for each variable (acceleration, gyroscope, magnetometer, free accelerations, cadence, velocity, AGV = accelerations, gyroscope and velocity, AGVC = accelerations, gyroscope, velocity and cadence) using four machine learning algorithms: K-Nearest Neighbors (KNN), Long Short-Term Memory (LSTM), Random Forest (RF), and Support Vector Machines (SVM). Magnetometer and cadence produced lower accuracies than combined variable trials for KNN, RF and SVM. There was no significant difference between the different variable combinations.

Table 4.43 Significant differences in mean accuracies for walking aid recognition cross-validation test between variables. Cad = Cadence, Mag = Magnetometer, Acc = Acceleration, FA = Free Acceleration

Model		Mean (%) ± *SD*		Mean (%) ± *SD*	Difference (%)	*p*-value
KNN	Mag	70.00 ± 1.27	AGV	76.35 ± 0.54	6.35	0.021
	Mag	70.00 ± 1.27	AGVC	76.67 ± 0.47	6.67	0.004
	Cad	65.66 ± 0.49	AGV	76.35 ± 0.54	10.69	0.002
	Cad	65.66 ± 0.49	AGVC	76.67 ± 0.47	11.01	<0.001
RF	Mag	72.90 ± 0.42	AGV	79.06 ± 0.47	6.16	0.021
	Mag	72.90 ± 0.42	AGVC	79.21 ± 0.83	6.31	0.001
	Cad	66.39 ± 0.76	AGV	79.06 ± 0.47	12.67	0.02
	Cad	66.39 ± 0.76	AGVC	79.21 ± 0.83	12.82	<0.001
	Cad	66.39 ± 0.76	All	78.71 ± 0.92	12.32	0.005
SVM	Mag	68.59 ± 0.99	AGV	83.51 ± 1.18	14.92	0.01
	Mag	68.59 ± 0.99	AGVC	84.16 ± 0.13	15.57	0.003
	Mag	68.59 ± 0.99	All	83.10 ± 0.42	14.51	0.04
	Cad	63.08 ± 0.52	AGV	83.51 ± 1.18	20.43	<0.001
	Cad	63.08 ± 0.52	AGVC	84.16 ± 0.13	21.08	<0.001
	Cad	63.08 ± 0.52	All	83.10 ± 0.42	20.12	0.004
	FA	73.87 ± 0.42	AGVC	84.16 ± 0.13	10.29	0.028
LR	Cad	63.05 ± 0.55	Acc	72.36 ± 0.99	9.31	0.004
	Cad	63.05 ± 0.55	AGV	73.23 ± 2.29	10.18	0.001
	Cad	63.05 ± 0.55	AGVC	71.72 ± 3.74	8.67	0.017
	Cad	63.05 ± 0.55	All	70.86 ± 1.34	7.81	0.032

Effects of normalisation on walking aid use recognition accuracies.

The data were normalised per stride and per person. To see if this is necessary, trials were run to see how the different normalisation methods and the not normalised raw data affect the classification accuracies. Normalising per stride produced smaller mean accuracies in eight out of the ten trials compared to raw and normalised per person data, so as with the terrain classification, a combined measure was run which only included raw and normalised per person data (RP). For leave-some-out trials there was no significant difference between the accuracies from Kruskal–Wallis tests (KNN $X^2(4)=1.72$ $p=0.79$, RF $X^2(4)=0.86$ $p=0.93$, SVM $X^2(4)=1.75$ $p=0.78$, LR $X^2(4)=1.70$ $p=0.79$, LSTM $X^2(4)=0.92$ $p=0.92$). For the cross-validation trials all modes showed statistical significance from Kruskal–Wallis tests (KNN $X^2(4)=18.47$ $p<0.001$, RF $X^2(4)=14.22$ $p=0.007$, SVM $X^2(4)=21.2$ $p<0.001$, LR $X^2(4)=9.93$ $p=0.042$, LSTM $X^2(4)=12.39$ $p=0.015$). Dunn's post-hoc tests found RP and all the normalisation methods combined produced similar accuracies for all models. For KNN, RF and SVM trials, RP and all the methods combined produced significantly higher accuracies than normalising per stride. RP was also significantly higher than raw data for KNN, and all combined had higher accuracies than Raw for SVM. Normalised per stride was significantly worse than normalising per person for LSTM. Results for significance testing are displayed in Table 4.46. These results disprove the null hypothesis that normalisation technique has no effect on walking aid recognition accuracies. As RP and all combined normalisation methods were statistically similar, RP will be used for the rest of this analysis.

Table 4.44 Leave-some-out walking aid recognition percentage accuracies for different normalisation techniques. RP = raw and person combined.

	KNN	RF	SVM	LR	LSTM
Raw	55.53%	60.46%	61.91%	57.58%	54.56%
Stride	56.41%	59.04%	58.28%	62.06%	55.11%
Person	54.93%	61.76%	62.54%	61.05%	57.37%
RP	57.47%	61.83%	63.18%	55.41%	60.11%
All	57.93%	61.36%	63.83%	57.80%	61.55%

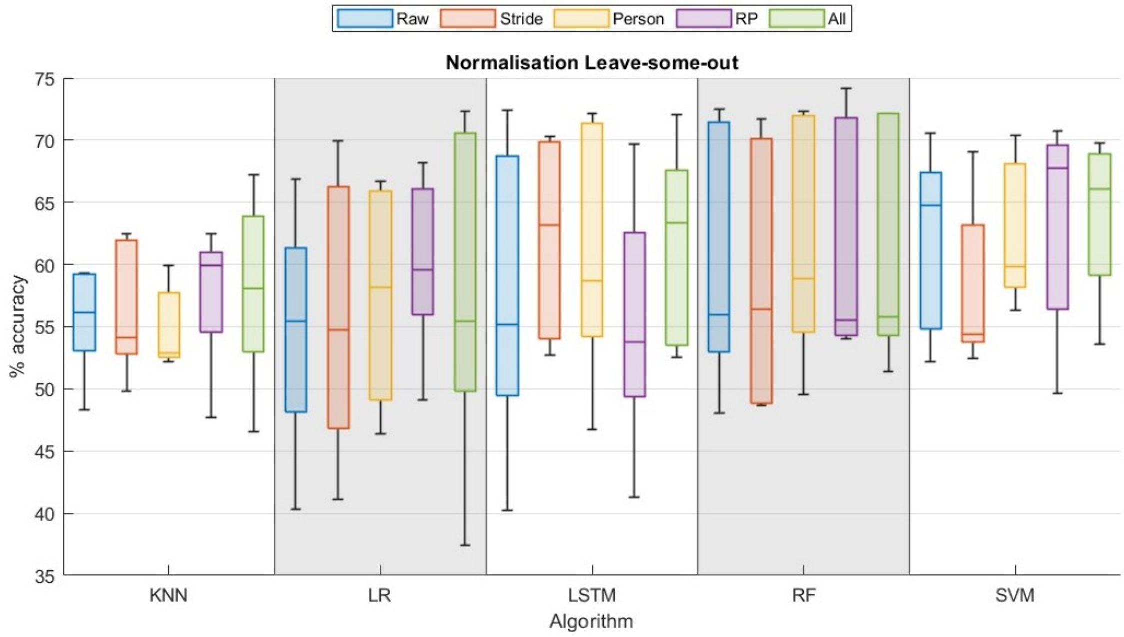

Fig. 4.36 Leave-some-out accuracy results for walking aid recognition across different normalisation methods. The boxplot shows the classification accuracy for each variable (raw data, normalised per stride, normalised per person, RP = raw and person combined and all data) using four machine learning algorithms: K-Nearest Neighbors (KNN), Long Short-Term Memory (LSTM), Random Forest (RF), and Support Vector Machines (SVM). There was no significant difference between the normalisation techniques.

Table 4.45 Cross-validation walking aid recognition percentage accuracies for different normalisation techniques. RP = raw and person combined.

	KNN	RF	SVM	LR	LSTM
Raw	74.30%	78.07%	78.92%	69.63%	65.87%
Stride	74.04%	76.31%	77.34%	68.97%	63.40%
Person	75.03%	78.29%	80.69%	70.75%	67.29%
RP	77.11%	79.63%	83.45%	72.70%	66.02%
All	76.67%	79.21%	84.16%	71.72%	65.65%

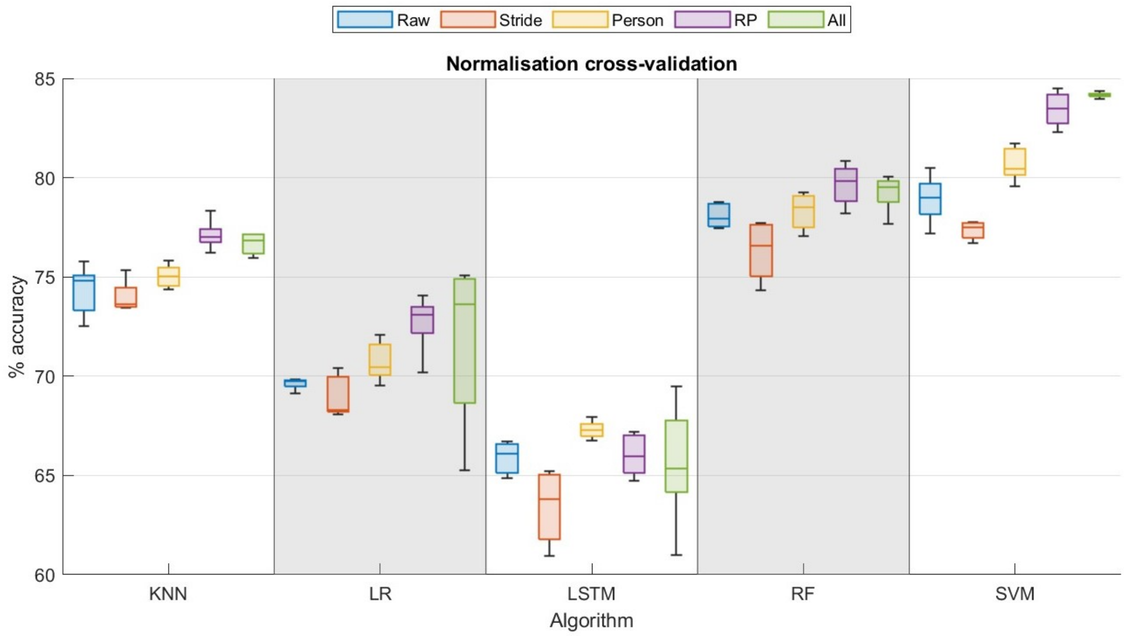

Fig. 4.37 Cross-validation accuracy results for walking aid recognition across different normalisation methods. The boxplot shows the classification accuracy for each variable (raw data, normalised per stride, normalised per person, RP = raw and person combined and all data) using four machine learning algorithms: K-Nearest Neighbors (KNN), Long Short-Term Memory (LSTM), Random Forest (RF), and Support Vector Machines (SVM). There was no significant difference between the normalisation techniques. RP and all the data combined produced higher accuracies than normalising per stride for KNN, RF and SVM. Normalizing per person had higher accuracies than normalising per stride for LSTM.

Table 4.46 Significant differences in mean accuracies for walking aid recognition cross-validation test between normalisation method. Per= normalised per person

Model		Mean (%) $\pm$ *SD*		Mean (%) $\pm$ *SD*	Difference (%)	*p*-value
KNN	Stride	74.01 $\pm$ 0.72	RP	77.11 $\pm$ 0.68	3.07	0.010
	Stride	74.01 $\pm$ 0.72	All	76.67 $\pm$ 0.47	2.63	0.024
	Raw	74.30 $\pm$ 1.13	RP	77.11 $\pm$ 0.68	2.81	0.037
RF	Stride	76.31 $\pm$ 1.33	RP	79.63 $\pm$ 0.94	3.32	0.010
	Stride	76.31 $\pm$ 1.33	All	79.21 $\pm$ 0.83	2.90	0.037
SVM	Stride	77.34 $\pm$ 0.41	RP	83.45 $\pm$ 0.80	6.11	0.008
	Stride	77.34 $\pm$ 0.41	All	84.16 $\pm$ 0.13	6.82	<0.001
	Raw	78.92 $\pm$ 1.09	All	84.16 $\pm$ 0.13	5.24	0.03
LSTM	Stride	63.40 $\pm$ 1.66	Per	67.29 $\pm$ 0.40	3.89	0.005

Effects of prosthetic type on walking aid use recognition accuracies.

As with the terrain classification, trials were run to see if splitting the data into prosthetic types would affect the accuracies of walking aid recognition. For the leave-some-out trials, the participants data were split into the same groups used in the previous analysis for terrain classification, section 4.3.2. Table 4.24 displays the grouping used for the trials. Mean accuracies for each of the leave-some-out and cross-validation trials are shown in Tables 4.47 and 4.48 and visualised in Figures 4.38 and 4.38. There was no significant improvement in the accuracies by splitting the data by prosthetic type compared to leaving the data combined for leave-some-out trials from Kruskal–Wallis tests (KNN $X^2(2)$=5.09 p=0.079, RF $X^2(2)$=4.38 p=0.11, SVM $X^2(2)$=3.38 p=0.18, LR $X^2(2)$=0.90 p=0.64, LSTM $X^2(2)$=1.20 p=0.55). For the cross-validation verification Kruskal–Wallis tests found significance for all models (KNN $X^2(2)$=11.2 p=0.004, RF $X^2(2)$=11.52 p=0.003, SVM $X^2(2)$=10.82 p=0.005, LR $X^2(2)$=6.03 p=0.049, LSTM $X^2(2)$=9.50 p=0.009). Dunn's post-hoc tests found the TF trials were significantly worse than the combined data for LSTM (TF M=57.64% SD=2.23, Combined M= 66.02% SD=0.96, difference=8.38% p=0.039), but TF was significantly better for LR (TF M=74.67% SD=2.93, Combined M=72.70% SD=1.32, difference=1.97% p=0.048). There was no significant difference between TT and combining all the data for any model this proves the null hypothesis that splitting the data by prosthetic type does not improve walking aid recognition accuracies. Hence, for this research, both prosthetic types will be combined.

Table 4.47 Leave-some-out walking aid recognition percentage accuracies for data split into prosthetic type. TT = Transtibial, TF = Transfemoral

	KNN	RF	SVM	LR	LSTM
TT	60.93%	70.47%	67.48%	48.54%	62.86%
TF	50.19%	52.11%	54.26%	51.32%	53.77%
Both	57.47%	61.83%	63.18%	55.41%	60.11%

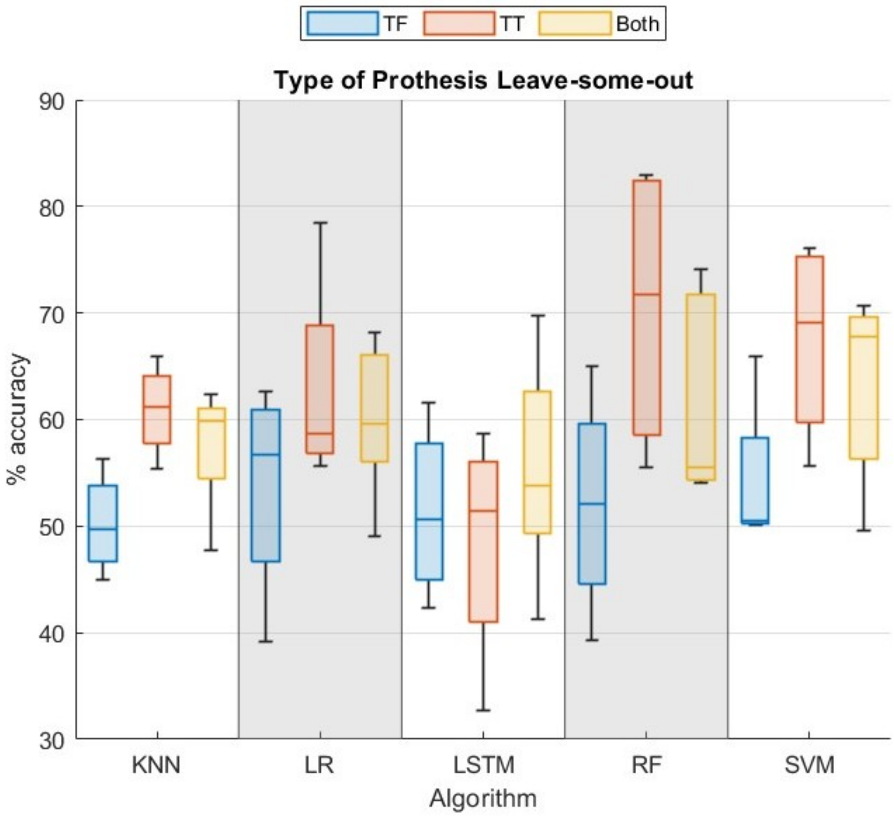

Fig. 4.38 Leave-some-out accuracy results for walking aid recognition when splitting different prosthetic types. The boxplot shows the classification accuracy for each prosthetic type (Transtibial-TT, Transfemoral-TF, both combined) using four machine learning algorithms: K-Nearest Neighbors (KNN), Long Short-Term Memory (LSTM), Random Forest (RF), and Support Vector Machines (SVM). There was no significant difference between the trials.

Table 4.48 Cross-validation walking aid recognition percentage accuracies for data split into prosthetic type. TT = Transtibial, TF = Transfemoral

	KNN	RF	SVM	LR	LSTM
TT	81.67%	78.55%	85.56%	74.66%	71.36%
TF	72.34%	77.33%	81.84%	74.67%	57.64%
Both	77.11%	79.63%	83.45%	72.70%	66.02%

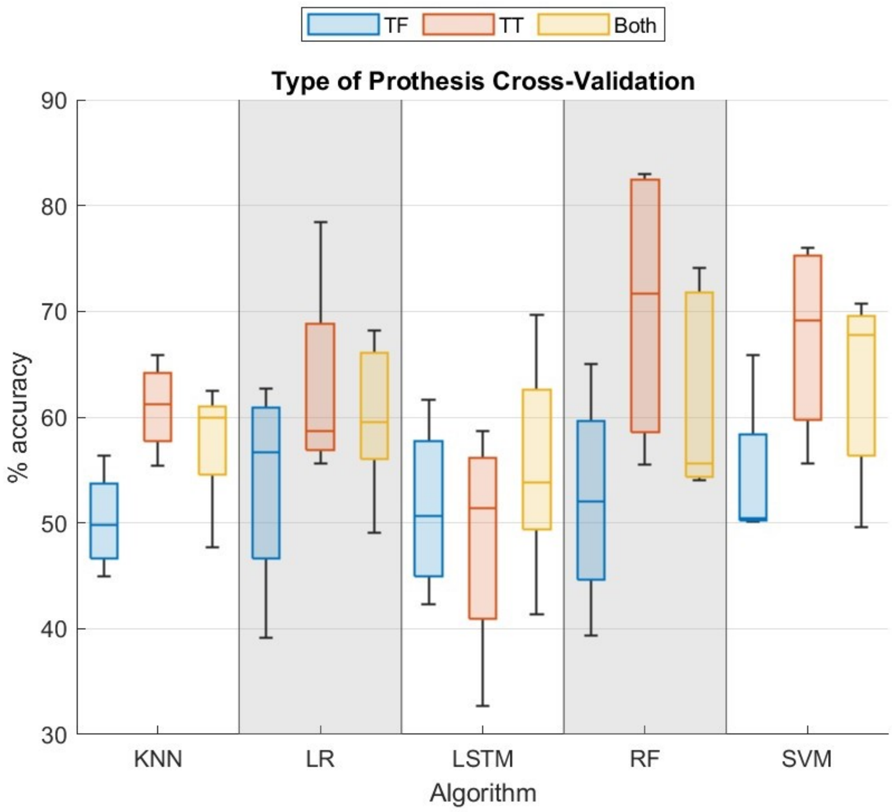

Fig. 4.39 Cross-validation accuracy results for walking aid recognition when splitting different prosthetic types. The boxplot shows the classification accuracy for each prosthetic type (Transtibial-TT, Transfemoral-TF, both combined) using four machine learning algorithms: K-Nearest Neighbors (KNN), Long Short-Term Memory (LSTM), Random Forest (RF), and Support Vector Machines (SVM). TF produced accuracies lower than the combined data for LSTM but higher for LR. There was no significant difference between TT and the combined data for any model.

Effects of terrain on walking aid use recognition accuracies.

Trials were run to see if the walking aid classification produced better accuracies with the data first split into the different terrains. As different participants walked a different number of strides on each terrain, and some did not have data collected on some terrains, only a cross-validation test was run for this part of the analysis. With cross-validation, all the strides are grouped together and split into equal groups, whereas if leave-some-out was to be used the number of strides in the test data could be close to the number used to train the classifiers and this could affect the accuracies. Kruskal–Wallis tests found statistical significance in all models (KNN $X^2(8)=26.98$ $p<0.001$, RF $X^2(8)=25.55$ $p=0.001$, SVM $X^2(8)=28.42$ $p<0.001$, LR $X^2(8)=34.51$ $p<0.001$, LSTM $X^2(8)=31.11$ $p<0.001$). Dunn's post-hoc tests found no individual terrain produced accuracies significantly higher than for the trials with all terrains combined. Due to this it would not be beneficial to split the data into terrains before classifying walking aid use which proves the null hypothesis.

Table 4.49 Cross-validation walking aid recognition percentage accuracies for different terrains.

	KNN	RF	SVM	LR	LSTM
Flat	64.28%	70.08%	74.11%	65.64%	55.34%
Grass	77.94%	80.32%	84.77%	79.03%	65.97%
Up Stairs	77.86%	79.00%	80.43%	69.43%	59.43%
Down Stairs	81.41%	83.10%	80.70%	70.28%	62.39%
Up Slope	79.34%	81.15%	86.05%	78.27%	68.90%
Down Slope	83.49%	83.69%	87.53%	78.85%	68.23%
Uneven	79.76%	81.73%	85.47%	80.67%	64.73%
Gravel	78.65%	80.02%	83.61%	74.52%	63.63%
All	77.11%	79.63%	83.45%	72.70%	66.02%

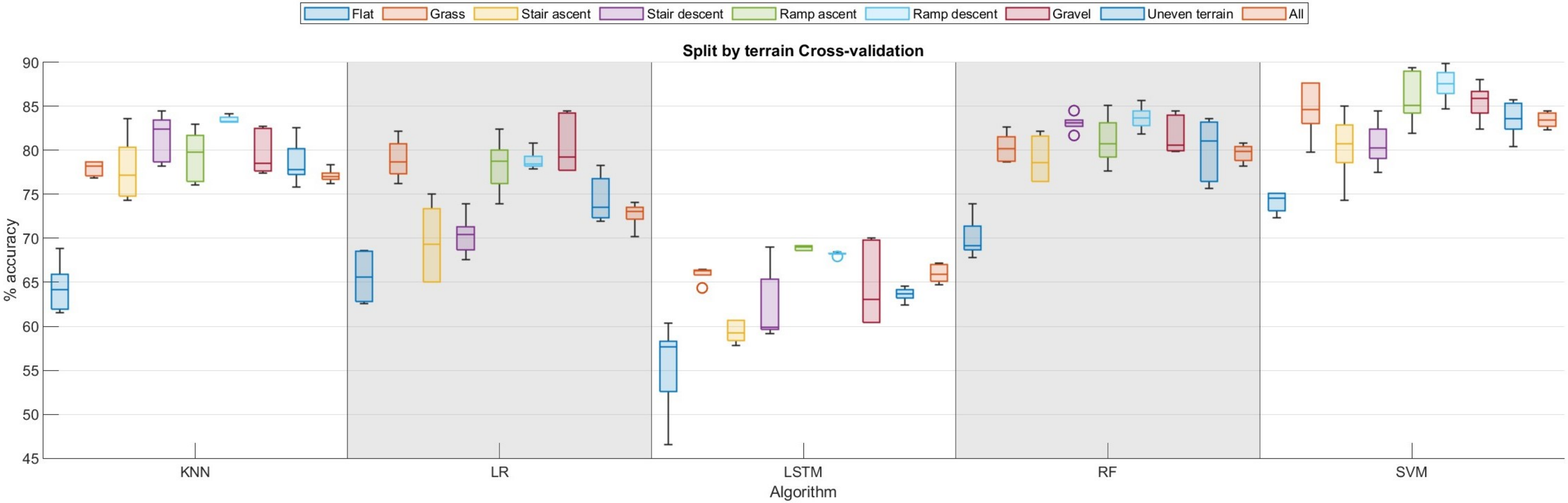

Fig. 4.40 Cross-validation accuracy results for walking aid recognition across different terrain. The boxplot shows the classification accuracy for each variable (Flat ground, Grass, Stair ascent, Stair descent, Ramp ascent, Ramp descent, Gravel, Uneven terrain and all combined) using four machine learning algorithms: K-Nearest Neighbors (KNN), Long Short-Term Memory (LSTM), Random Forest (RF), and Support Vector Machines (SVM). No trial with a single terrain isolated produced accuracies significantly higher than the trials with all the terrains combined.

Final system informed by previous findings for walking aid use recognition.

The following summarizes the findings of the walking aid use classification :

- An IMU on the trunk is the best single IMU to identify walking aid use, and using a combination of the four IMUs produced the best accuracies, but there was not a large difference between this and a single IMU on the prosthetic shank.
- Splitting the data into windows that contain a whole stride produced the best accuracies.
- Reducing the data to 10 datapoints did not reduce the accuracies of the classifiers.
- Acceleration, gyroscope, velocities and cadence combined produced the best accuracies, whereas magnetometer and free accelerations did not improve the accuracies.
- A combination of data normalised per person and raw data produce the best accuracies.
- There could be benefits to analysing TT participants separately but not TF participants.
- There is no benefit in classifying data for separate terrains.
- RF and SVM produced better accuracies than LR, LSTM and KNN.

An optimiser was run on RF and SVM classifiers with just the accelerations, gyroscope, velocity and cadence data, split into stride-based window of 10 datapoints, with raw data and data normalised per person, to see if adjusting the hyperparameters would improve the classification accuracies. The hyperparameters for the RF model were: number of trees – 350, number of predictors – 78, minimum leaf size – 1. The hyperparameters for the SVM model were: box constraint – 915.93, kernel score – 12.962, and using one vs one coding. The optimised classifiers were then compared to the previous classifiers. Table 4.50 and 4.51 display the mean accuracies and Figures 4.41 and 4.42 visualises the results for leave-some-out and cross-validation. There was no significant difference between the optimised classifiers and the previous ones for any of the classifiers and verification tests from Kruskal–Wallis tests (Leave-some-out - RF $X^2(1)=0.27$ $p=0.60$, SVM $X^2(1)=0.53$ $p=0.46$, cross-validation - RF $X^2(1)=1.84$ $p=0.17$, SVM $X^2(1)=0.27$ $p=0.60$). SVM produced a better accuracy for both leave-some-out and cross-validation. The optimised SVM classifier did produce the highest accuracies for both leave-some-out and cross-validation.

Table 4.50 Leave-some-out walking aid recognition percentage accuracies for trial with 10 datapoints using the previous models and the models with optimised hyperparameters.

	RF	SVM
previous	61.83%	63.18%
Optimised	59.91%	65.03%

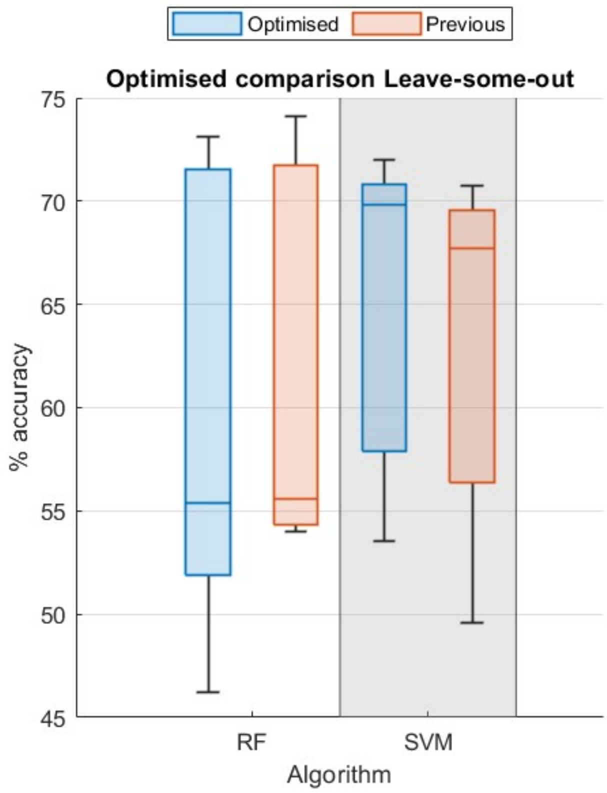

Fig. 4.41 Leave-some-out accuracy results for walking aid recognition across trials with 10 datapoints using models with the hyperparameters that have been used for all of the analyses and models that have had their hyperparameters optimised. The boxplot shows the classification accuracy for the two types of models (Previous -using hyperparameters used in previous analysis, Optimised – using optimised hyperparameters) using the best performing machine learning algorithms from the previous analysis: Random Forest (RF), and Support Vector Machines (SVM). There is no significant difference between the trials.

Table 4.51 Leave-some-out walking aid recognition percentage accuracies for trial with 10 datapoints using the previous models and the models with optimised hyperparameters.

	RF	SVM
Previous	79.63%	83.45%
Optimised	80.40%	84.42%

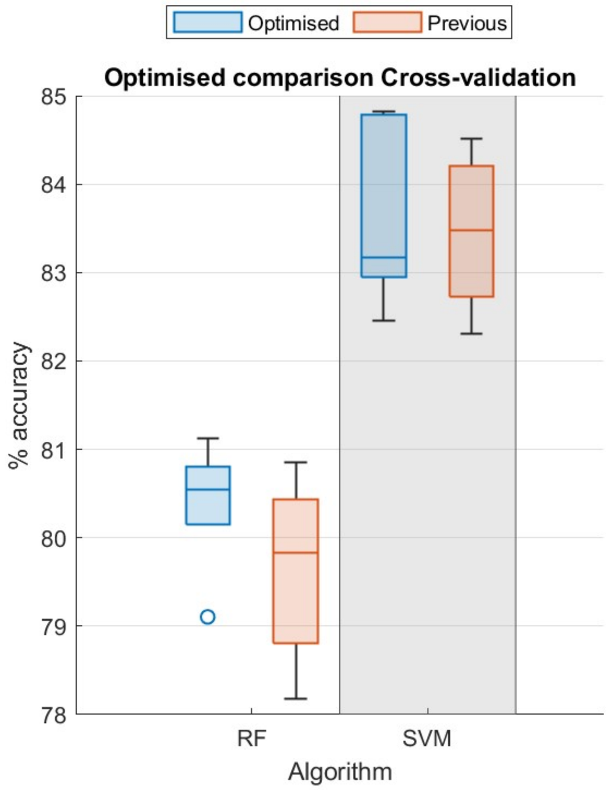

Fig. 4.42 Cross-validation accuracy results for walking aid recognition across trials with 10 datapoints using models with the hyperparameters that have been used for all of the analysis and models that have had their hyperparameters optimised. The boxplot shows the classification accuracy for the two types of models (Previous -using hyperparameters used in previous analysis, Optimised – using optimised hyperparameters) using the best performing machine learning algorithms from the previous analysis: Random Forest (RF), and Support Vector Machines (SVM). The optimised trials produced higher mean accuracies but not to a significant level.

4.4 Discussion and Conclusions

4.4.1 Terrain classification

For terrain classification there remain questions on how different terrains are classified. One of these questions is at what degree should ground be classified as a slope. According to UK building regulations a 1:20 incline is considered a slope [176]. For this research, all slopes were steeper then 1:20. On all the slopes used outside the motion capture laboratory there were variations in their steepness. This could have affected the accuracies and could have increased gait variations. This could make the model more robust and improve the accuracies for real-world use, as the model was trained on more variations of slope walking. How to define uneven terrain is another question. For this research, cobble stones and uneven pavement were used as uneven terrain, but there was no set requirement as to how uneven the terrain had to be to be labelled as uneven; it was decided by the researcher conducting the study. Even with terrain that was considered very uneven there could be places where the terrain is flat but as the whole section was labelled uneven these strides would be considered uneven. This could be a source of lower accuracies when classifying uneven terrain as seen in the confusion matrixes Figures 4.43 and 4.44.

Figure 4.43 is the confusion matrix for the leave-some-out trial for the optimised RF model. As can be seen, the model slightly over-classified strides as flat, and this could have been because there were more flat strides than for any other terrain. Most of this overclassification comes from grass, uneven and gravel. Stair ascending and descending produced the best accuracies. Jamieson [40] found that accuracies increased when grouping flat terrain with uneven, gravel and grass as one group, and the results from this research would suggest the same. The cross-validation confusion matrix, Figure 4.44, shows some of these issues but also seems to show some problems with classifying stair ascending. Jamieson [40] achieved a leave-some-out accuracy of 56.68% and a cross-validation accuracy of 78.46%, whereas for the model created in this research using the same terrain groupings achieved a leave-some-out accuracy of 79.02% and a cross-validation accuracy of 90.29%. These accuracies are much higher but as can be seen in Figure 4.45 and 4.46 the model overfits for flat terrain due to the disproportion of strides in the flat terrain group. There are methods that could counter this overfitting, weighting the classes to force the model to prioritise different classes, running the optimisers to adjust the model or collecting more data on stairs and ramps to balance the stride in each class grouping [177].

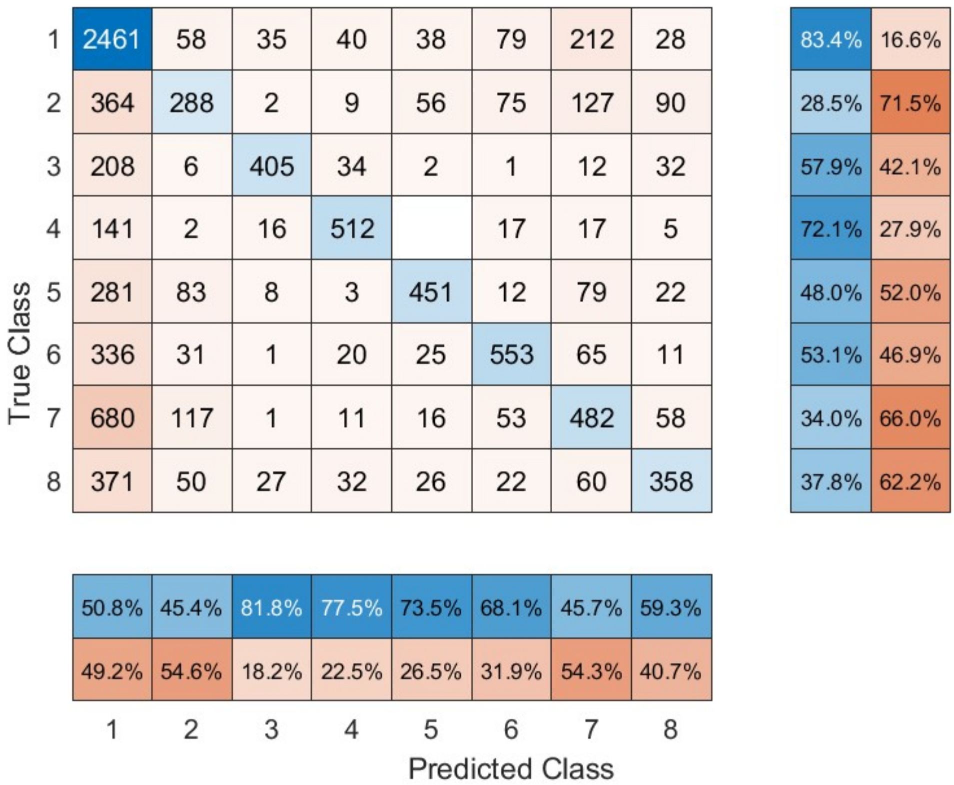

Fig. 4.43 Confusion matrix for terrain classification RF leave-some-out 1-Flat 2-Grass 3-Stair ascent 4-Stairs descent 5-Ramp ascent 6-Ramp descent 7-Uneven terrain 8-Gravel.

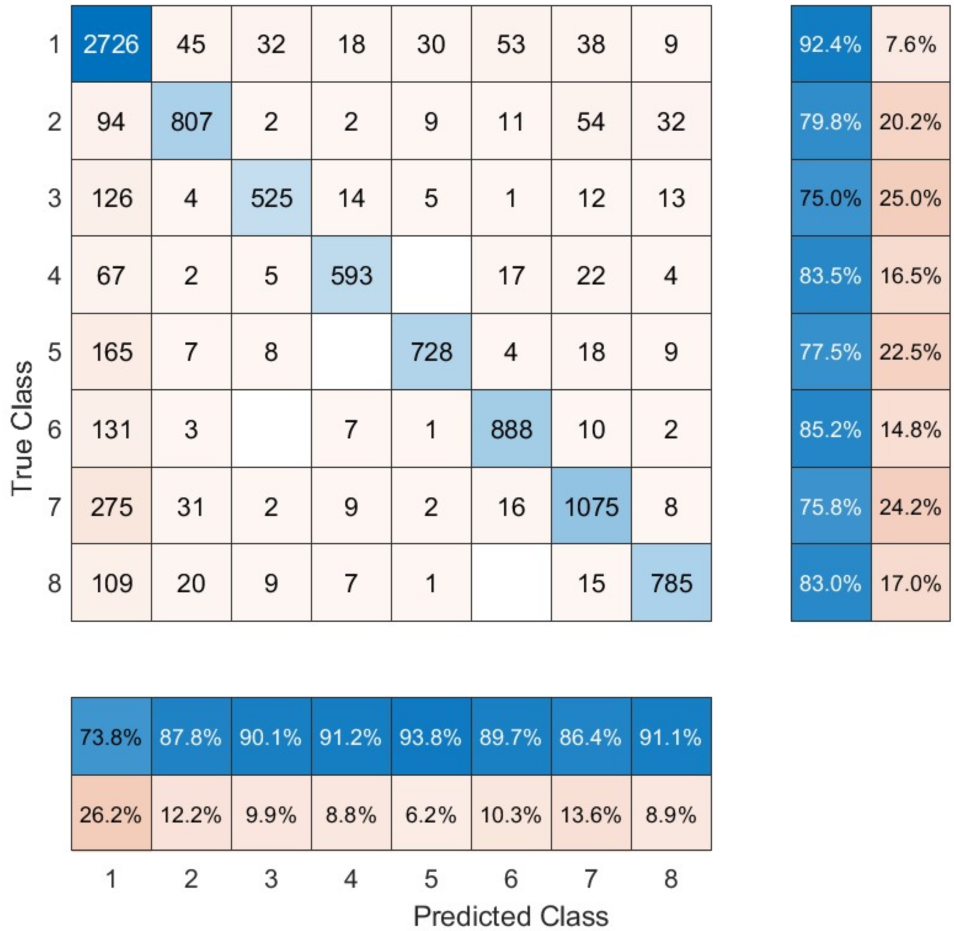

Fig. 4.44 Confusion matrix for terrain classification RF Cross-validation 1-Flat 2-Grass 3-Stair ascent 4-Stairs descent 5-Ramp ascent 6-Ramp descent 7-Uneven terrain 8-Gravel.

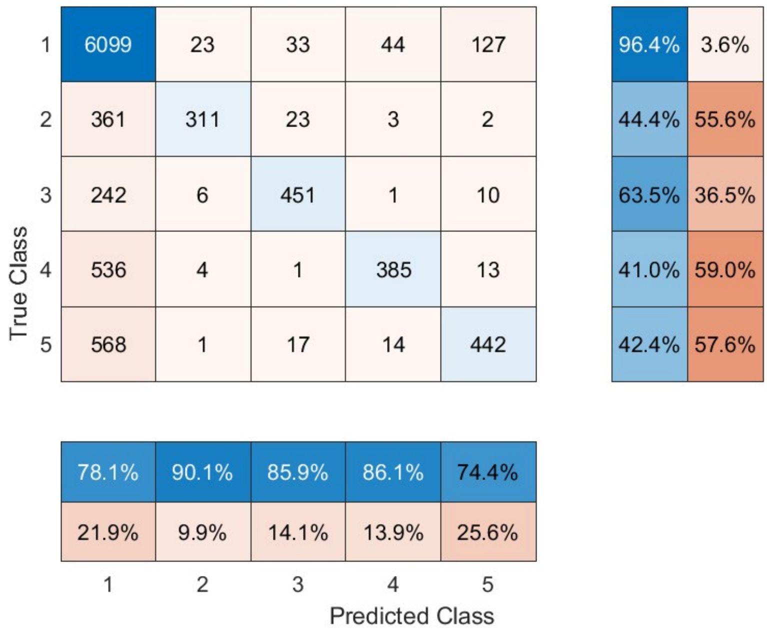

Fig. 4.45 Confusion matrix for terrain classification RF leave-some-out 1-Flat 2-Stair ascent 3-Stair descent 4- Ramp ascent, 5-Ramp descent.

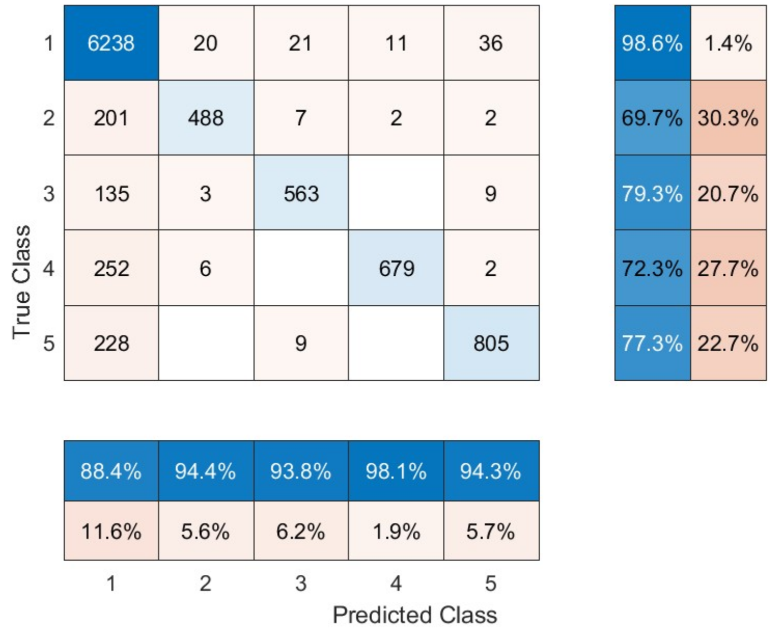

Fig. 4.46 Confusion matrix for terrain classification RF Cross-validation 1-Flat 2-Stair ascent 3-Stair descent 4- Ramp ascent, 5-Ramp descent.

Another question is how to deal with transition steps. In this research transitions steps were not identified. Some studies have found transitions steps to be different [120][107][121], and therefore they could affect accuracies if not labelled correctly. There is also the question of when a stride should be classified as a certain terrain. For this research as the stride is counted during the swing phase, the terrain the foot lands on during the stride is labelled but no analysis was done to identify which parts of the gait are the most important for the classifiers and therefore where the transition between different gait patterns for different terrains occurs. Transition steps would eliminate this by giving them a different label.

4.4.2 Walking aid use classification

Sixty-one percent of the strides recorded were taken without a walking aid. This means the leave-some-out accuracies achieved by the walking aid use recognition models were only 4.03% higher than the percentage split in strides between strides with a walking aid and without. The cross-validation accuracies were much higher but still had 16.33% misclassification. If a patient only used a walking aid for a proportion of their daily strides or didn't use one at all but the classifier says they use one for 16.33% of their day, this could change the view a clinician has on this patient and therefore affect their K level assessment. Walking aids are used to increase balance and offload [178][179]. If offloading could be measured, by measuring foot pressure or load through the prosthesis, this combined with the IMU data could have potential to produce a better classification accuracy. Figure 4.47 shows the confusion matrix for the leave-some-out trial of the optimised SVM model. The model is overclassifying to without a walking aid which could be due to the number of strides with a walking aid being 61% of all the steps. Class weighting techniques could be used to adjust for this overclassification but if the research was to continue and collect more data with a walking aid this could also help fix this issue. The cross-validation confusion matrix, Figure 4.48, also shows signs of this issue.

Fig. 4.47 Confusion matrix for walking aid recognition Leave-some-out SVM optimised trial. 1-withouth a walking aid 2-with a walking aid.

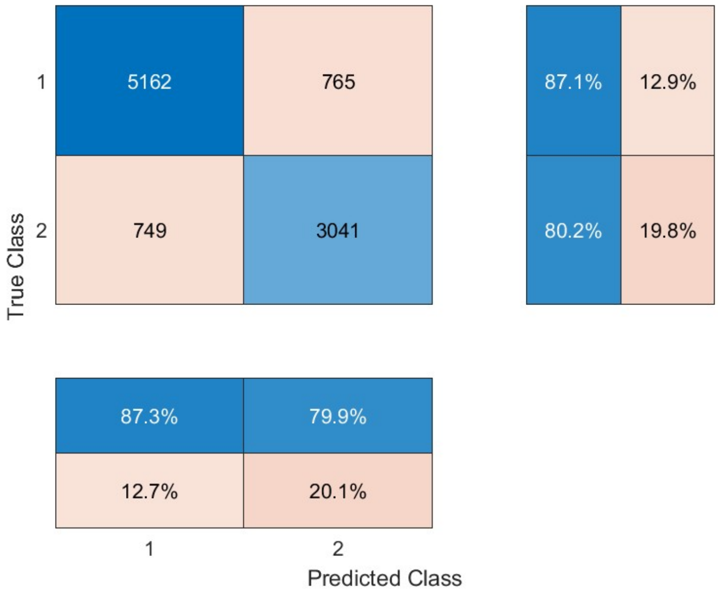

Fig. 4.48 Confusion matrix for walking aid recognition cross-validation SVM optimised trial. 1-withouth a walking aid 2-with a walking aid..

One question that arises in this research was what to class as a walking aid. This research counted handrails as a walking aid as it was deemed that a participant would use it to offload and therefore could change their gait.

The only personal walking aid used in the research was a walking stick. This was the only personal walking aid the participants were comfortable using and used in their everyday life. There is currently no information on what are the most commonly used walking aids for lower limb prosthetic users. If available, the participants were asked to use their own walking stick but if it was not available a walking stick was provided. Due to this, the classifier has not been trained on other walking aids, such as walking frames or rollators, so it cannot be assumed it would be able to classify their use. If it is deemed that other walking aids would need to be classified; further research would have to capture data of participants using these walking aids.

For this research sliding windows were not assessed. In previous research sliding windows have shown to be effective [96], but there has been no consensus on the best window style and if windows should be sliding. Sliding windows in their nature do produce additional datapoints as the same data is used in multiple windows, but this is not unique information. For this research it was deemed that comparing time-based windows to a stride-based window was more important. Further research could be done to compare sliding time-based windows to a stride-based window, but at present the stride-based window produced the best accuracies.

The demographics of the participants did not perfectly match the estimated demographics of prosthetic users in the UK[12]. Only 10% of the participants were female. However, about 30% of lower limb prosthetic users are female in the United Kingdom. The range of age was 33 to 85 years with 85% above the age of 54. In comparison, 70% of lower limb prosthetic users in the United Kingdom are above 54. However, 54% of lower limb prosthetic users in the United Kingdom are TT and 37% TF, with 94% being unilateral, which are closely matched in this research with 55% TT, 40% TF and 95% unilateral. The demographics of the study do not perfectly match the demographics of lower limb prosthetic users in the United Kingdom, especially for sex, but come close in relation to unilateral/bilateral and TT/TF. For this research it was decided that, instead of restricting participant number to meet the demographics of prosthetic users in the UK, collecting as much data as possible was more important. The 20 participants recruited for this research provided a larger sample than for any similar study found in Chapter 3, with the largest being 9 transtibial by Du, L [130]. All the participants conducted the study in just 2 areas in the country, Greater Manchester and South Hampshire, and this means that the classifiers are only trained on terrain in these areas and therefore may not be able to classify terrain and variations in terrain that occur in different parts of the country.

The leave-some-out analysis was carried out in five groups of participants to reduce computational time compared to individual leave-one-out analysis. This allowed for more aspects of the analysis to be investigated, for example different windowing methods and effects of splitting the data by prosthesis type. This did reduce the classification accuracies, as shown in section 4.3.2, but the comparative accuracies are still valid. The cross-validation analysis was conducted as 5-fold, whereas some previous studies have used 10-fold [109][106][112]. This may have increased sensitivity to anomalies and variability in results. However, the reduced computational time enabled a more extensive analysis. The largest variance was observed in the isolated velocity LSTM model (mean accuracy: 72.64%, variance: 4.84), but this variance is not substantial.

In conclusion;

The cross-validation and leave-some-out accuracies achieved in the terrain recognition were higher than achieved by a similar study that used a thigh mounted accelerometer to classify the terrain a lower limb prosthetic user was traversing [40]. Collecting more data could help to increase the leave-some-out accuracy and get it closer to the cross-validation accuracy. This research has proven that it is possible to classify the terrain a lower limb prosthetic user is traversing using a prosthetic shank mounted IMU, but more data, from more participants in different locations, would have to be collected to improve the accuracies and make the classifiers more robust.

The walking aid recognition showed issues of overclassification. This could be due to the percentage difference in strides with a walking aid as compared to without that were used to train the classifier. Collecting more data of walking aid use could help reduce this issue. Alternatively, additional sensors that could measure the amount of offloading in combination with the IMU could help to increase the accuracies.

This chapter designed a first system comprised of sensors and algorithms for classification of terrain and walking aid use using real-world data. The next chapter (Chapter 5) will utilize in-lab 3D motion capture to record lower limb prosthetic users traversing different terrains with and without a walking aid, to then use the data to create virtual IMUs, to see if sensor placement on the limb would affect the accuracy of the classification algorithms.

Chapter 5

System refinement: Simulating virtual sensors from stereophotogrammetry data to explore effects of sensor position on activity classification accuracies.

5.1 Background

In the previous chapter (Chapter 4), a first sensor system and associated algorithms for activity classification were designed. The purpose of this chapter was to refine the system through use of simulated, virtual sensors, and thereby improve the accuracies of the machine learning classification models created in Chapter 4.

Machine learning classification works best with data with high variance between classes but low variance within classes [180]. During the data collection for Chapter 4, the location of the IMU on the prosthetic shank was not consistent as it was set where it was convenient for the participant. Hence sometimes the IMU was placed closer to the ankle and sometimes closer to the knee, and for some participants it was moved slightly along the shank during the study because the original location proved problematic for the participant. This could have caused variability in the data as the acceleration pattern changes along the shank for a stride. The best IMU location is where the signal variance is highest between different terrains and walking aid use conditions, but lowest for the same terrain or walking aid use condition. To calculate where this position would be, virtual sensors can be created from stereophotogrammetry data [181][182].

System refinement: Simulating virtual sensors from stereophotogrammetry data to explore effects of sensor position on activity classification accuracies.

Studies have previously found that small changes in IMU position can create high errors in different measures [183][184][185]. Angular velocities should be consistent along a limb, as long as the IMU is placed with the same orientation and on the same plane. Magnetometer data, which measures orientation, should also be consistent along a limb if the IMU is placed with the same orientation. Creating virtual accelerometers from stereophotogrammetry data has previously been done in work by Tresadern, et al. [182] and Tong and Granat [181]. The same method as Tresadern, et al. [182] will be used in this research.

5.1.1 Aims of Chapter 5

- To collect 3D motion capture stereophotogrammetry data alongside sensors data inside the gait lab for lower limb prosthetic users transversing different terrains , providing a comprehensive dataset for analysing sensor placement effects.
- To simulate virtual sensor data in different positions along the prosthetic limb and assess the effects of sensor placement on classification accuracies for activity and terrain types.

5.2 Methodology

In this work, stereophotogrammetry data were collected in a motion capture laboratory that could then be used to simulate virtual sensor data. Data collection took place at the University of Salford where a Qualisys motion capture system was used, with the cameras calibrated before each participant so that the position error was less than 1mm. Data were collected for traversing four different types of terrain, including flat ground, stairs, ramp and uneven terrain, and these terrain data were obtained for walking with and without a walking aid. Ethical approval was granted to collect the required gait lab data (Ethical approval numbers for University Ethics: 4743 Appendix D.2).

The following Methods sections describe the different walking conditions that data were collected for, the participants, and data collection aspects, as well as virtual sensor design.

5.2.1 Terrain

Flat Ground

Flat ground walking trials took place on the smooth, vinyl flooring of the motion capture laboratory. The capture area the participants walked in was 5.0m long and 2.0m wide.

Stairs

The set of stairs used for this research had three steps on one side and four on the other, with a flat section on the top. This gave some variation in the step size and allowed both stair assending and deceding to occur in the same trial, reducing the number of trials the participants were asked to do. All the steps were 260mm in depth. The three steps on the one side were 200mm high, and the four steps on the other side were 150mm high. 150mm is the minimum height regulated in the UK for a set of stairs [176]. 220mm is the maximum height for private stairs and 190mm for utility stairs [176], to have the higher stairs within this range, and to ensure the stairs are the same height wihtin the 600mm height, 200mm was chosen as the higher stair height. The flat section on the top was 700mm long. The set of stairs had a uniform width of 620mm. There was also an adjustable hand rail that was set at a comfortable height for the participants. Figure 5.1 displays the dimensions of the stairs. Each participant was asked to complete four trials on the stairs with and without a walking aid.

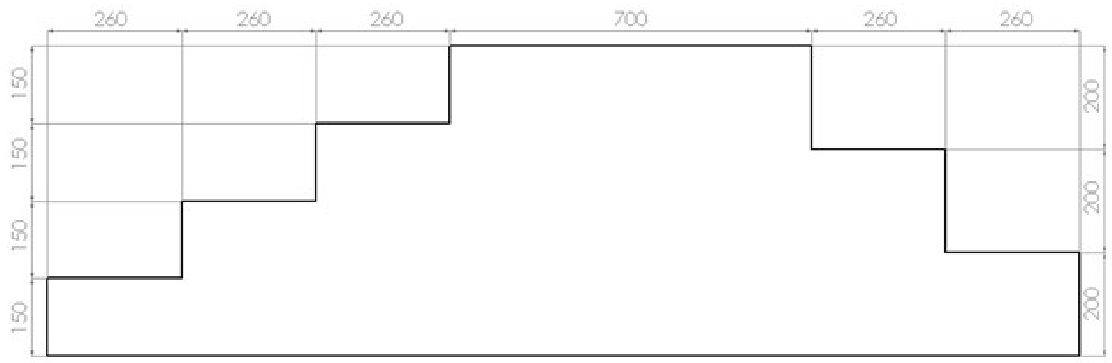

Fig. 5.1 Set of stairs used for data collection in the gait lab. Lengths in mm.

Ramp

The ramp used in the research was 1500mm long and 130mm high to give a constant gradient of 4.95°, which is larger than the 1:20 (2.9°) deemed a slope [176]. There was a flat section at the top of the ramp that was 1500mm long. The ramp and flat section were both 1500mm wide. Participants did separate ascending and descending trials, to eliminate the participant turning 180° during a trial. Due to this, each participant was asked to complete 4 ascending and descending trials with and without a walking aid.

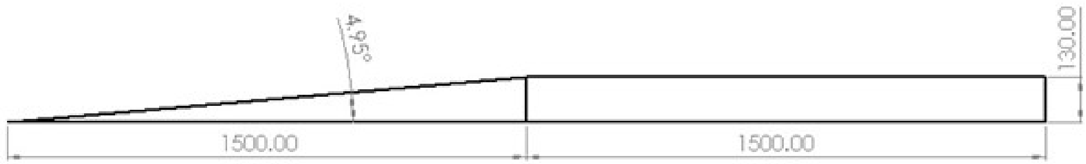

Fig. 5.2 Ramp set up as used for data collection in the gait lab. Lengths in mm.

Uneven Ground

To simulate uneven terrain, artificial cobblestones were produced that could be used in the motion capture laboratory. Twenty of the cobblestones that were traversed as part of the data collection at the University of Salford campus in Chapter 4 were randomly selected, and their height, width, length and their distance to the neighbouring cobble stones were measured. The maximum and minimum for each of these measurements were used to design individual cobblestones for the gait lab set up, thereby recreating a realistic replica of a cobblestone path for this indoor study.

Table 5.1 Cobblestone dimensions

Measurement	Measurement range
Length	120mm – 175mm
Width	60mm – 90mm
Height	12mm – 19mm
Distance between stones	40mm – 55mm

Specifically, the cobblestones were designed by randomly allocating a measurement within the range measured for height, width, and length. To increase the unevenness of the terrain, it was decided to allow the distance between the cobblestones to be increased. The cobblestones were designed on a 200mm by 125mm rectangle, with the cobblestone placed randomly at least 10mm from one of the 200mm and 125mm edges. This ensured the minimum distance between the cobblestones to be 10mm but the largest distance could be 150mm and 120mm. The cobblestones were then randomly ordered into a 5 by 16 grid, i.e. a total of 80 cobblestones made up the cobblestone pathway which was 2000mm by 1000mm in its overall dimensions. Figure 5.3 visualises the the artificial cobblestones. With an average healthy stride length of 780mm [186], this guaranteed that at least one stride was taking off and landing on the cobble stones for each foot. The largest distances between the cobblestones when put into the gird were 129mm and

101mm, and the smallest distances were 15mm and 16mm. Appendix H.1 displays the dimensions for each cobble stone. The cobblestones were machined out of medium-density fibreboard.

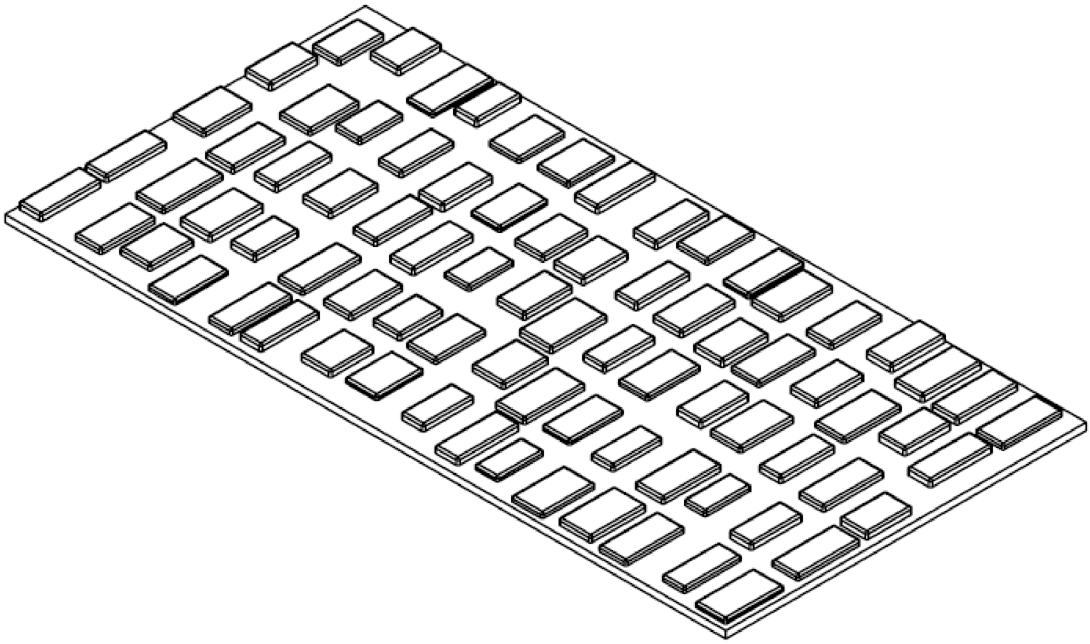

Fig. 5.3 Artificial cobblestones

5.2.2 Walking aid use

The only personal walking aid used by participants taking part in the research was a walking stick. This was the only personal walking aid the participants were comfortable using and used in their everyday life. If available, the participants were asked to use their own walking stick but if it was not available a walking stick was provided.

5.2.3 Participants

This study used a subset of participants that had also been used in the algorithm creation study (Chapter 4), and they had to meet the same inclusion and exclusion criteria as detailed in Chapter 4. There were ten participants who were all unilateral prosthetic users and which provided informed consent. All the participants were male. Six of the participants were transtibial and four were transfemoral. The age range was 53 to 85 years old with 90% above 54. Table 5.2 displays the details for each participant.
Three of the participants did not feel comfortable walking up or down the stairs with a walking aid, hence they did not complete that part of the study. All other trials were completed by all participants. Table 5.3 displays the trials performed by each participant.

Table 5.2 Participant Information.

Participant no.	Participant no. from Chapter 4	Age	Sex	Transtibial (TT)/ Transfemoral (TF)
1	1	85	Male	TT
2	2	63	Male	TT
3	3	69	Male	TF
4	5	74	Male	TT
5	6	69	Male	TT
6	8	56	Male	TT
7	9	53	Male	TF
8	10	59	Male	TF
9	11	72	Male	TT
10	12	64	Male	TF

Table 5.3 Trials completed by each participant. 'wi': with walking aid, 'wo': without walking aid.

Participant no.	Flat	Up Stairs	Down Stairs	Up Slope	Down Slope	Uneven
1	wi/wo	wo	wo	wi/wo	wi/wo	wi/wo
2	wi/wo	wi/wo	wi/wo	wi/wo	wi/wo	wi/wo
3	wi/wo	wi/wo	wi/wo	wi/wo	wi/wo	wi/wo
4	wi/wo	wi/wo	wi/wo	wi/wo	wi/wo	wi/wo
5	wi/wo	wi/wo	wi/wo	wi/wo	wi/wo	wi/wo
6	wi/wo	wi/wo	wi/wo	wi/wo	wi/wo	wi/wo
7	wi/wo	wi/wo	wi/wo	wi/wo	wi/wo	wi/wo
8	wi/wo	wo	wo	wi/wo	wi/wo	wi/wo
9	wi/wo	wi/wo	wi/wo	wi/wo	wi/wo	wi/wo
10	wi/wo	wo	wo	wi/wo	wi/wo	wi/wo

5.2.4 Procedure

Marker Placement

The motion capture marker placement used a six degrees of freedom model, as shown in Figure 5.4. The six degrees of freedom model was chosen because it has shown to produce smaller errors than the conventional gait model and does not require excessive maker placements, so is not off-putting for the participants, but fully identifies the locations this research is focusing on [187].

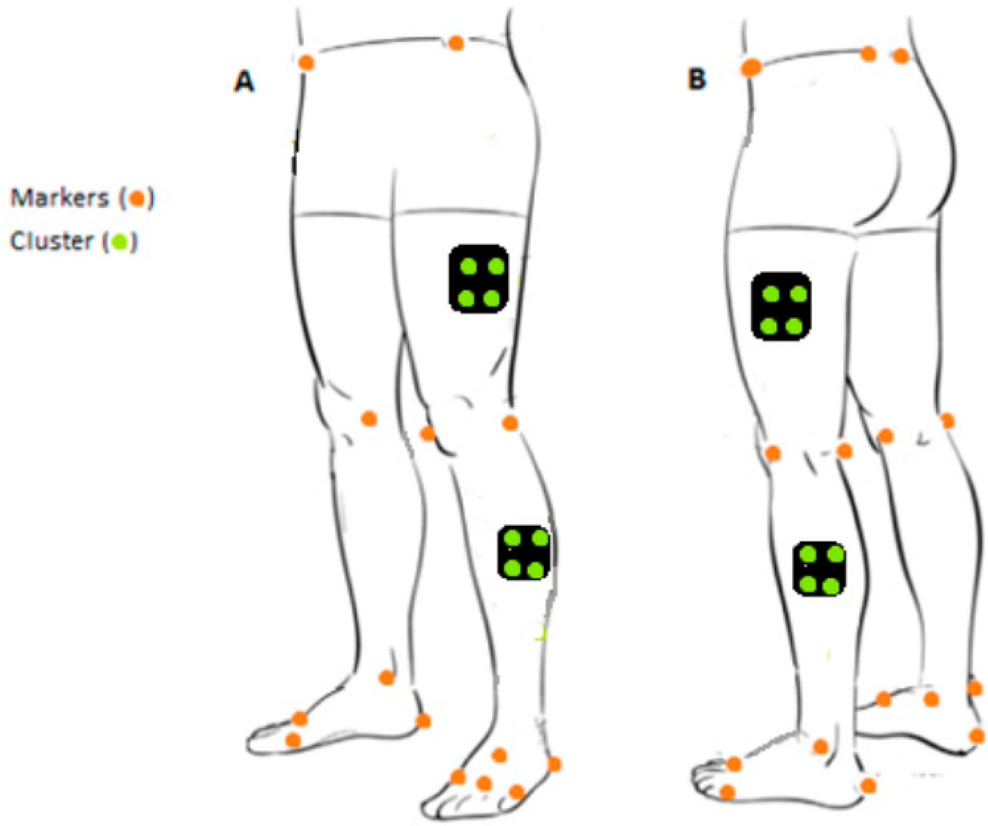

Fig. 5.4 Six degrees of freedom gait model.

Trial Procedure

The procedure for the study was the same as detailed in Chapter 4 (section 4.2.5). After establishing that the participant was comfortable conducting the trial, the procedure was as follows:

- Participant confirms they are ready.
- Start motion capture.
- Start IMU recording.
- Give the participant the signal to start.
- Participant stamps their prosthetic leg on the floor twice.
- The participant conducts the trial.
- The participant finishes the trial and stands still.
- IMU recording is stopped.

- Motion capture recording is stopped.

As in Chapter 4, the participants were asked to stamp their prosthetic leg twice to make it easier to align the video and IMU data.

5.2.5 Analysis

Feature Importance

The first stage of the analysis was to investigate which aspects of the accelerations are key for the classification models from Chapter 4. Then it was investigated whether the variation of these features of the accelerations are increased when the IMU is placed at different locations. To identify what features of the accelerations had the biggest influence on the classification models, three different techniques were used with the outcomes scored and collated to rank the features based on their influence on the classification outcome.

Minimum Redundancy Maximum Relevance (mRMR)

Minimum Redundancy Maximum Relevance, mRMR is a feature selection technique used to identify features with high relevance but low redundancy [188]. Higher relevance means features that are more aligned to the target class, and low redundancy means features that are less correlated to other features. The mRMR function used for this project was the built-in Matlab mRMR function that returns a score for each feature, with higher scores given to more important features, i.e. features that have high relevance and low redundancy [189].

Neighborhood component analysis (NCA)

Neighborhood component analysis, NCA was developed by J. Goldberger et al. at the University of Toronto [190]). This analysis aims to learn a low-dimensional representation of data such that the classification accuracy is maximized. The algorithm learns a linear transformation of the feature space, allowing it to identify the most informative features while reducing the dimensionality. Similar to KNN algorithms, a NCA algorithm tries to minimise the distance between data points of the same class while maximising the distance between different classes. The built-in Matlab function for NCA was used for this project. The feature weights output of this function was used to determine the most influential features which have larger weights [191].

Recursive Feature Elimination (RFE)

Recursive Feature Elimination, RFE ranks features by their importance based on how much they contribute to the model's predictions. The main idea behind RFE is to recursively remove the least important features of a model until the optimal subset of features is found [192]. RFE was used for the KNN, SVM and RF algorithms for terrain classification, and RF and SVM for walking aid recognition to identify the features that are most important to each algorithm.

Scoring

The three methods, mRMR, NCA and RFE, were run on the reduced featured data set determined in Chapter 4. This included the accelerations, gyroscope, magnetometer, free accelerations and cadence data for terrain classification, and accelerations, gyroscope, velocity and cadence data for walking aid recognition, with data normalised per participant and raw data. Table 5.4 displays the features included in both analyses. As the data set is time-dependant, the mean, median and maximum for each feature importance method were calculated for each feature. RFE was also run on leave-some-out and cross-validation data. To reduce the computational need, the data were run in strides with 10 data points per stride, which from Chapter 4 did not significantly reduce the accuracies of the algorithms. There were 24 values for feature importance for terrain classification and 18 for walking aid recognition, as described in Table 5.4. For three methods each features will be ranked with 1 being the most important and the least important feature having the heights rank. Then the total rankings will be summed up and the overall most important features will be determined, as having the lowest overall score. As the gyroscope and orientation measures do not change with IMU placement, only the acceleration data were analysed.

Table 5.4 Features used for terrain and walking aid classification.

Feature	Reference name	Terrain / walking aid classification
Acceleration in the vertical direction (w/gravity)	X	Both
Acceleration in the medio-lateral direction (w/gravity)	Y	Both
Acceleration in the anterior-posterior direction (w/gravity)	Z	Both
Acceleration in the vertical direction (w/o gravity)	Free X	Terrain
Acceleration in the medio-lateral direction (w/o gravity)	Free Y	Terrain

Acceleration in the anterior-posterior direction (w/o gravity)	Free Z	Terrain
Rate of turn along the vertical direction	Gyroscope X	Both
Rate of turn along the medio-lateral direction	Gyroscope Y	Both
Rate of turn along the anteriorposterior direction	Gyroscope Z	Both
3D magnetic field in the vertical direction	Magnetometer X	Terrain
3D magnetic field in the medio-lateral direction	Magnetometer Y	Terrain
3D magnetic field in the anteriorposterior direction	Magnetometer Z	Terrain
Delta_velocity (dv) in the vertical direction	Velocity X	Walking aid
Delta_velocity (dv) in the mediolateral direction	Velocity Y	Walking aid
Delta_velocity (dv) in the anteriorposterior direction	Velocity Z	Walking aid
Resultant accelerations (w/gravity)	Resultant	Both
Resultant accelerations (w/o gravity)	Free Resultant	Terrain
Number of strides per second	Cadence	Both
Normalised per person acceleration in the vertical direction (w/gravity)	NPP X	Both
Normalised per person acceleration in the medio-lateral direction (w/gravity)	NPP Y	Both
Normalised per person acceleration in the anterior-posterior direction (w/gravity)	NPP Z	Both
Normalised per person acceleration in the vertical direction (w/o gravity)	NPP Free X	Terrain
Normalised per person acceleration in the medio-lateral direction (w/o gravity)	NPP Free Y	Terrain
Normalised per person acceleration in the anterior-posterior direction (w/o gravity)	NPP Free Z	Terrain
Normalised per person rate of turn along the vertical direction	NPP Gyroscope X	Both
Normalised per person rate of turn along the medio-lateral direction	NPP Gyroscope Y	Both

Normalised per person rate of turn along the anteriorposterior direction	NPP Gyroscope Z	Both
Normalised per person 3D magnetic field in the vertical direction	NPP Magnetometer X	Terrain
Normalised per person 3D magnetic field in the medio-lateral direction	NPP Magnetometer Y	Terrain
Normalised per person 3D magnetic field in the anteriorposterior direction	NPP Magnetometer Z	Terrain
Normalised per person delta_velocity (dv) in the vertical direction	NPP Velocity X	Walking aid
Normalised per person delta_velocity (dv) in the mediolateral direction	NPP Velocity Y	Walking aid
Normalised per person delta_velocity (dv) in the anteriorposterior direction	NPP Velocity Z	Walking aid
Normalised per person resultant accelerations (w/gravity)	NPP resultant	Both
Normalised per person resultant accelerations (w/o gravity)	NPP Free resultant	Terrain
Normalised per person number of strides per second	NPP Cadence	Both

5.2.6 Creating virtual sensors

The process of creating the virtual accelerometer data was taken from Tresadern, et al. [182]. The process involved creating a reference coordinate frame and inferring its position and orientation with respect to the world reference frame. The world reference frame is the reference frame created by the Qualisys system in the motion capture laboratory, and this is fixed for all trials. For this research only the prosthetic shank was investigated. It was assumed that the prosthetic shank is a rigid body. Therefore, the reference frame for this study was taken to be between the medial and lateral malleolus makers and the medial and lateral epicondyle makers for the prosthetic leg. The dynamics of the reference coordinate frame are calculated by finite differencing from the position and orientation. To investigate how the acceleration changes along the shank, accelerations were calculated at three positions. The three positions were mid-shank, towards the knee and towards the ankle. The positions towards the knee and towards the ankle were taken at 5% of the shank length away from the knee and ankle so were 90% of the shank length away from each other. These positions were chosen as it would be impractical to position an IMU directly at the knee or ankle joint. All the virtual sensors were placed on the anterior of the shank. With the virtual accelerations created, the same 5Hz second order Butterworth filter was used as was in Chapter 4, to smoothen the data.

To check the accuracy of the virtual sensor algorithm, accelerations were calculated at the location of a real IMU and the signals compared. A pilot study with a healthy adult was carried out with the participant walking on flat ground. The IMU was placed on the shank cluster and the four markers of the cluster used as the reference frame. It was assumed that the plane of the IMU and cluster were the same, and that the IMU records accelerations from the centre of the casing. The cross-correlation between the resultant acceleration was calculated and the results displayed in Figure 5.5, this method was also used by Tong et al. [181]. The highest peak for the cross-correlation is at zero which indicates no phase shift between the signals. Pearson correlation coefficient was calculated for the three acceleration components and resultant acceleration against the real IMU signals to obtain a general understanding of the overall similarity. Although this technique has not previously been conducted on acceleration data it has on time series data [193]. Table 5.5 displays the calculated coefficients and Figures 5.6 and 5.7 display the comparison between the real and virtual accelerations. The coefficients show great similarities between the X, Z, and resultant accelerations. There is a slightly lower value in the Y accelerations but it is still a strong correlation so was deemed acceptable for this study.

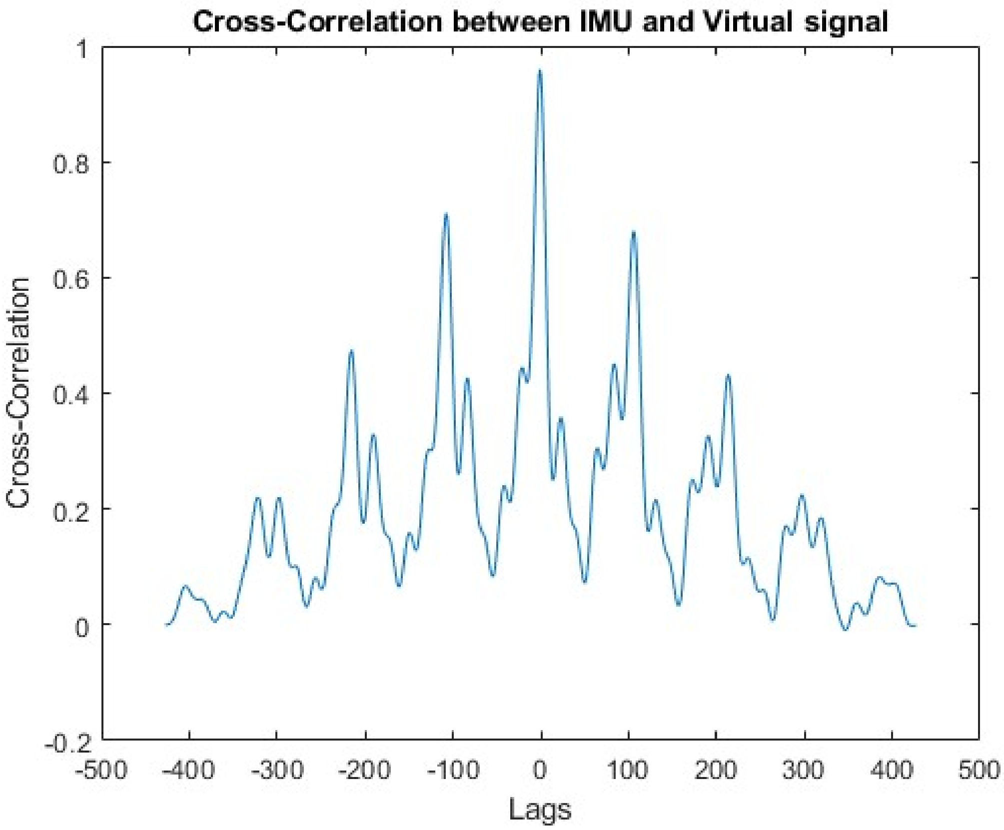

Fig. 5.5 Cross-correlation between IM and virtual signal for resultant acceleration. The highest peak is at 0 which indicates no phase shift between the signals

Table 5.5 Pearson correlation coefficients for comparison of real and simulated sensor data.

Acceleration	Pearson correlation coefficient
X accelerations	0.9658
Y accelerations	0.8725
Z accelerations	0.9320
Resultant accelerations	0.9532

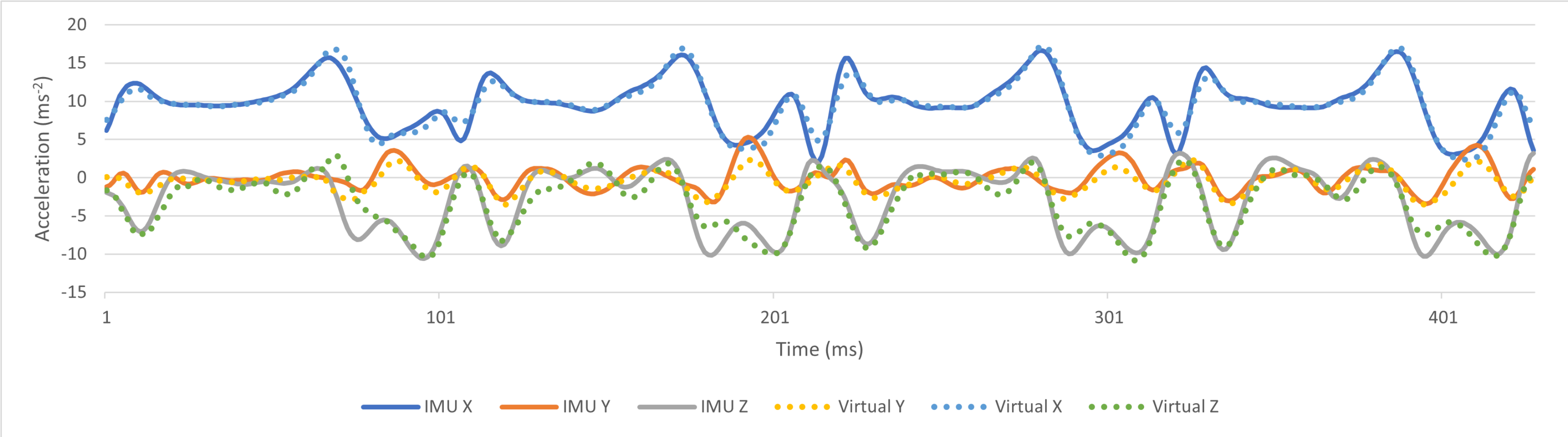

Fig. 5.6 Real and virtual acceleration comparisons for all 3 axis. There is strong correlation between the virtual and real accelerations.

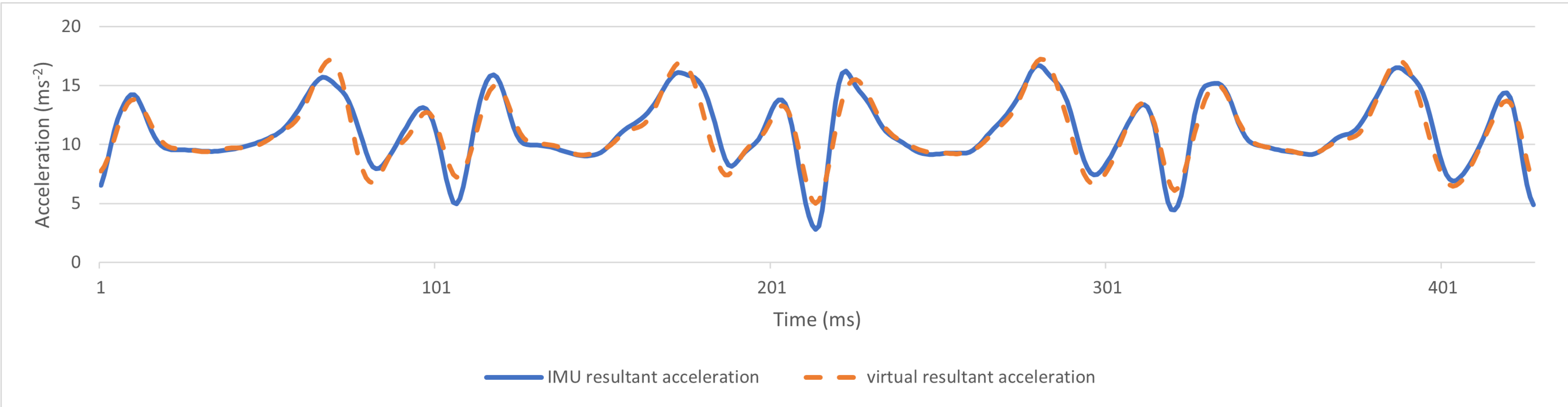

Fig. 5.7 Real and virtual acceleration comparisons of the resultant acceleration. There is strong correlation between the virtual and real accelerations

5.2.7 Stride count method

For this research only the acceleration data were processed. This means the same method of identifying and splitting the data into strides as Chapter 4, using the Y component of the gyroscope data, could not be used. Instead, the resultant accelerations of the prosthetic shank were used. As found in Chapter 4 section 4.2.7 this method had a 11.54% error in stride count and a 10.93% error in precision. To combat this error, each trial was manually checked and stride count and locations where a stride is counted manually changed if needed. This was possible due to the relatively small number of strides being assessed (585 strides). Figure 5.8 shows the resultant acceleration, Y component of the gyroscope data and force plate reading for a flat walking trial. The step count was always taken on the first spike of the resultant acceleration data which can be seen is in a similar location as the spike in the Y compound of the gyroscope data, which is in the swing phase of the stride.

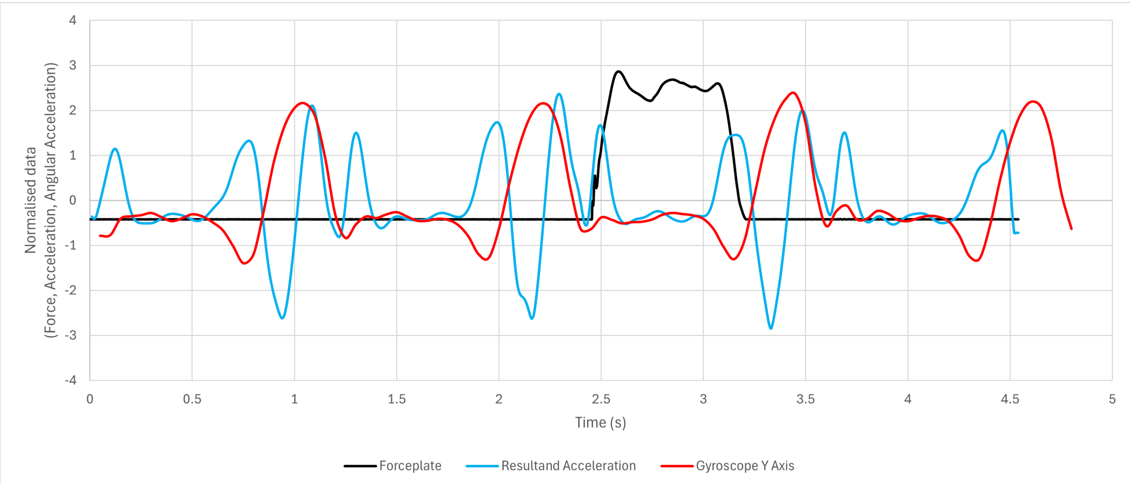

Fig. 5.8 Stride count comparison. Gyroscope Y axis (red) was used in Chapter 4 and strides counted at the peaks, Resultant Accelerations (light blue) was used in the study document in this chapter and peaks counted at the first peak after the minimum. Force plate (black) is included to indicate where heel strike and toe off is for the second stride.

5.2.8 Assessing variance between types of terrain to identify promising IMU position(s) for classification

Larger variance between conditions, but low variance within, is vital for successful activity classification. To assess the variance between acceleration signals for each terrain pair and each virtual IMU location, Kruskal-Wallis tests were run on the data to produce p values for each terrain comparison pair for the three IMU placements. Kruskal-Wallis tests were run because some of the data were not normally distributed. The Kruskal-Wallis tests were run for all features at one hundred stride segments. The p values were calculated for each feature at each time segment. The smallest p values will then be investigated, looking at the feature and time segment to see if the plot of the acceleration feature shows large variance at that time point. This was done for each of the three virtual IMU placements.

In addition, the Cohen's d, which is a standardised mean difference (SMD) method where the mean difference between two datasets is divided by the pooled standard deviation [194], was also calculated. SMD is a simple technique that can be used alongside Kruskal-Wallis to check the results make logical sense. As with the Kruskal-Wallis assessment, SMD was run for comparison of every terrain pair. The maximum SMD was found with the feature and percentage of a stride where it was produced.

5.2.9 Machine learning

As a final test to see how IMU placement affects the classification, the data for the three simulated IMUs were used to train and test the classification models developed in Chapter 4. The number of useable strides each participant performed on each terrain is displayed in Table 5.6. Due to the low number of strides and the variability between the number of strides that each participant performed on the different terrains, 5-fold cross-validation was chosen as the assessment method for the classification results. The percentage accuracy was calculated the same way as in Chapter 4.

Table 5.6 Number of strides each participant performed on each terrain.

Participant	Flat	Up Stairs	Down Stairs	Up Ramp	Down Ramp	Uneven
1	15	6	8	8	13	10
2	11	14	13	4	4	4
3	21	20	19	4	4	13
4	18	16	16	5	7	9
5	23	21	18	5	6	13
6	4	10	10	4	3	4
7	15	17	13	5	6	9
8	8	4	4	4	4	2
9	11	13	14	3	6	3
10	14	11	10	3	5	15

The data from three IMUs were also combined and used to train and test the classification models. This was done to see whether having different IMU locations in the same dataset affects the accuracies of the classifiers.

5.3 Results

5.3.1 Terrain classification

Results for the terrain classification analysis are presented first. As stated earlier on, this analysis looked firstly at which features affected the accuracies of the terrain classification models created in Chapter 4 most, and if there were differences in these features for the three simulated IMU positions. Then the analysis investigated how variance of the accelerations changes between each terrain comparison, and how this variance changes between the three simulated IMU locations. The final analysis then used the simulated accelerations to train and test terrain classification modes to investigate how the accuracies are affected by the different IMU locations, and also how combining data from IMUs in different locations could affect the accuracies.

Feature importance Three feature importance methods, mRMR, NCA and REF, were run on the data collected in Chapter 4. The features were then scored based on the results of each analysis, as described in section 5.2.5. The scored results for all the measures for feature importance in terrain classification of the real-world data are presented in Table 5.7. The total score, average score and median score, is displayed for each feature.

Table 5.7 Feature importance for terrain classification. Acceleration feature highlighted in grey. Total score, average score, and median score is displayed for each feature

Features	Total	Average	Median
NPP Free Z	222	9.25	7
Gyroscope Y	233	9.71	8
NPP Free Y	245	10.21	9.5
NPP Gyroscope Y	265	11.04	11
NPP Magnetometer Z	289	12.04	14
NPP Z	290	12.08	10.5
NPP Magnetometer X	293	12.21	8.5
NPP X	306	12.75	12.5
NPP resultant	312	13.00	13
X	316	13.17	13
Free Z	343	14.29	14
NPP Free resultant	348	14.50	13.5
Free Y	352	14.67	13
Magnetometer X	363	15.13	18
Y	376	15.67	18
NPP Y	379	15.79	15
Magnetometer Z	386	16.08	18
Z	393	16.38	19
Gyroscope X	394	16.42	17
Resultant	408	17.00	16.5
NPP Gyroscope Z	411	17.13	16
Gyroscope Z	414	17.25	17.5
NPP Magnetometer Y	415	17.29	18.5
NPP Cadence	430	17.92	20
Magnetometer Y	443	18.46	21
Free resultant	459	19.13	22
NPP Free X	464	19.33	21.5
NPP Gyroscope X	522	21.75	25.5
Free X	529	22.04	24
Cadence	560	23.33	26.5

Table 5.8 presents the acceleration measures from Table 5.9. It should be noted that features can score lower because they are similar to another feature, so are deemed to be more redundant and not providing unique information.

Table 5.8 Acceleration scores from Table 5.7 Total score, average score, and median score is displayed for each feature.

Features	Total	Average	Median
NPP Free Z	222	9.25	7
NPP Free Y	245	10.21	9.5
NPP Z	290	12.08	10.5
NPP X	306	12.75	12.5
NPP resultant	312	13.00	13
X	316	13.17	13
Free Z	343	14.29	14
NPP Free resultant	348	14.50	13.5
Free Y	352	14.67	13
Y	376	15.67	18
NPP Y	379	15.79	15
Z	393	16.38	19
Resultant	408	17.00	16.5
Free resultant	459	19.13	22
NPP Free X	464	19.33	21.5
Free X	529	22.04	24

For Table 5.8 it can be seen that accelerations in X, Y and Z planes are relevant for the classification accuracy. Table 5.8 suggests that normalised per person accelerations are more influential to the classification accuracy than non-normalised accelerations. The most influential features for individual acceleration axis and resultant accelerations were examined. Plots of the mean normalised per person free acceleration in the Z and Y axis and normalised per person acceleration in the X axis and NPP resultant accelerations against proportion of a stride are presented for the simulated accelerations in Figures 5.7 to 5.10.

It can be seen in Figure 5.9, normalised per person free accelerations in the Z axis, there is clear variance between stairs ascending and descending to the other terrains at 0.2 of a stride for the ankle and mid-shank. The variation for stair use to the other terrains for the knee is more apparent towards the end of the stride, where there is also variation for flat terrain.

Figure 5.10 shows, that normalised per person free accelerations in the Y axis, for the ankle and mid-shank, did not produce a smooth plot;, this could be because the values are relatively small. At the knee there is clear variation between stair ascending and descending to the other terrains, at between 0.8 and 0.9 of the stride. Stair ascending also has variance to the other terrains at the end of the stride.

For normalised per person accelerations in the X axis, Figure 5.11, there is variation for stair descending to the other terrains for all placements at about 0.8 of the stride;, the variation is lower for the knee. The ankle also has clear variation for stair ascending and descending to the other terrains at about 0.15 of a stride.

Figure 5.12 shows that, for normalised per person resultant accelerations, there is clear variation between stair descending and the other terrain at about 0.8 of the stride for all locations. There are also variations between flat and the other terrains at this point for the ankle and mid-shank, with 0.2 of the stride being another place for this variation for the ankle.

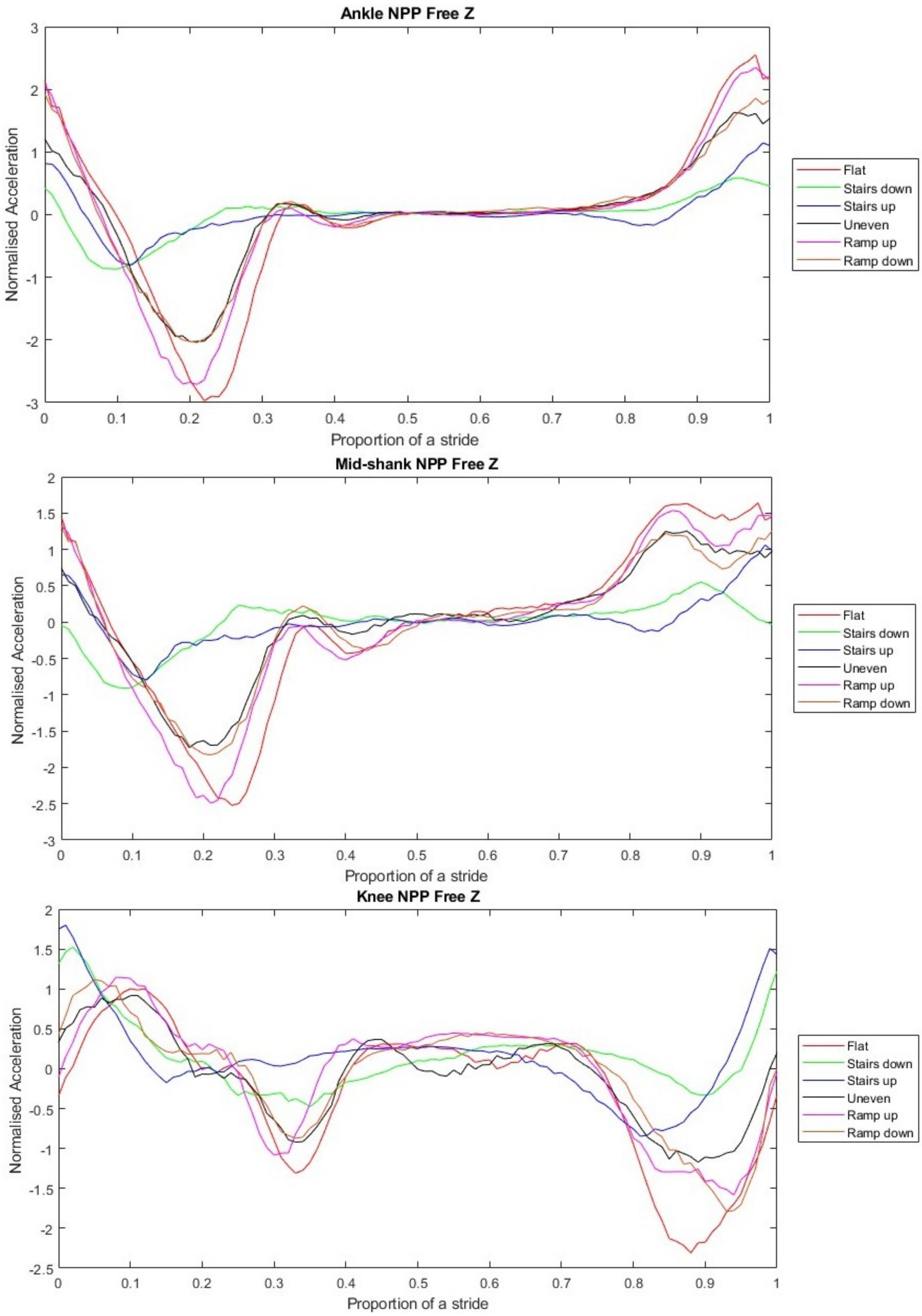

Fig. 5.9 NPP free Z for ankle (top), mid shank (middle), and knee (bottom). For the normalised per person free accelerations in the Z axis there is clear variance between stairs ascending and descending to the other terrains at 0.2 of a stride for the ankle (top) and mid-shank (middle). The variation for the knee (bottom) is more apparent towards the end of the stride, where there is also variation for flat terrain.

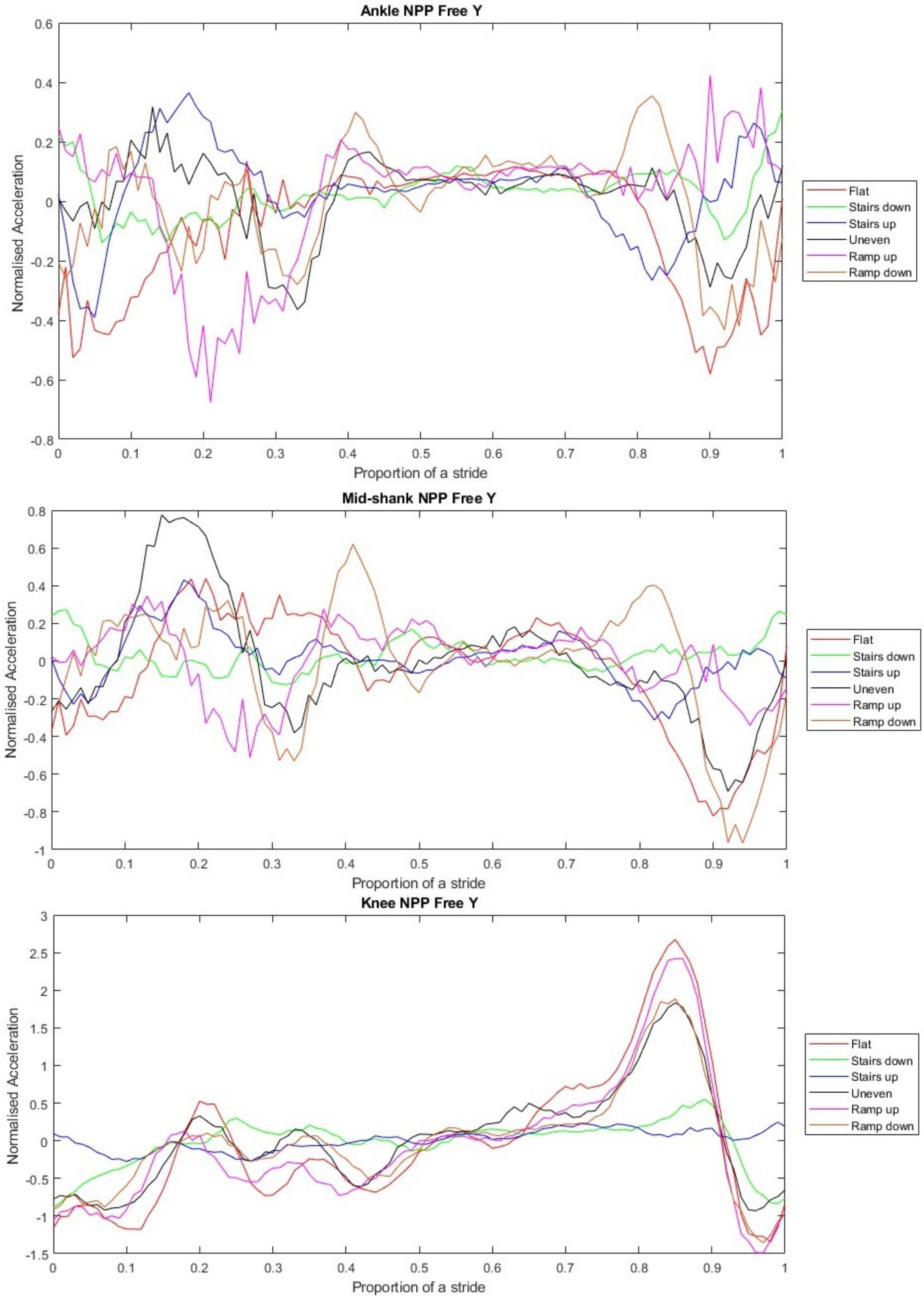

Fig. 5.10 NPP free Y for ankle (top), mid shank (middle), and knee (bottom). The mean normalised per person free accelerations in the Y axis for the ankle (top) and mid-shank (middle) did not produce smooth plots, this could be because the values were relatively small. At the knee (bottom) there was clear variation between stair ascending and descending to the other terrains, at between 0.8 and 0.9 of the stride. Stair ascending had variance to the other terrains at the end of the stride.

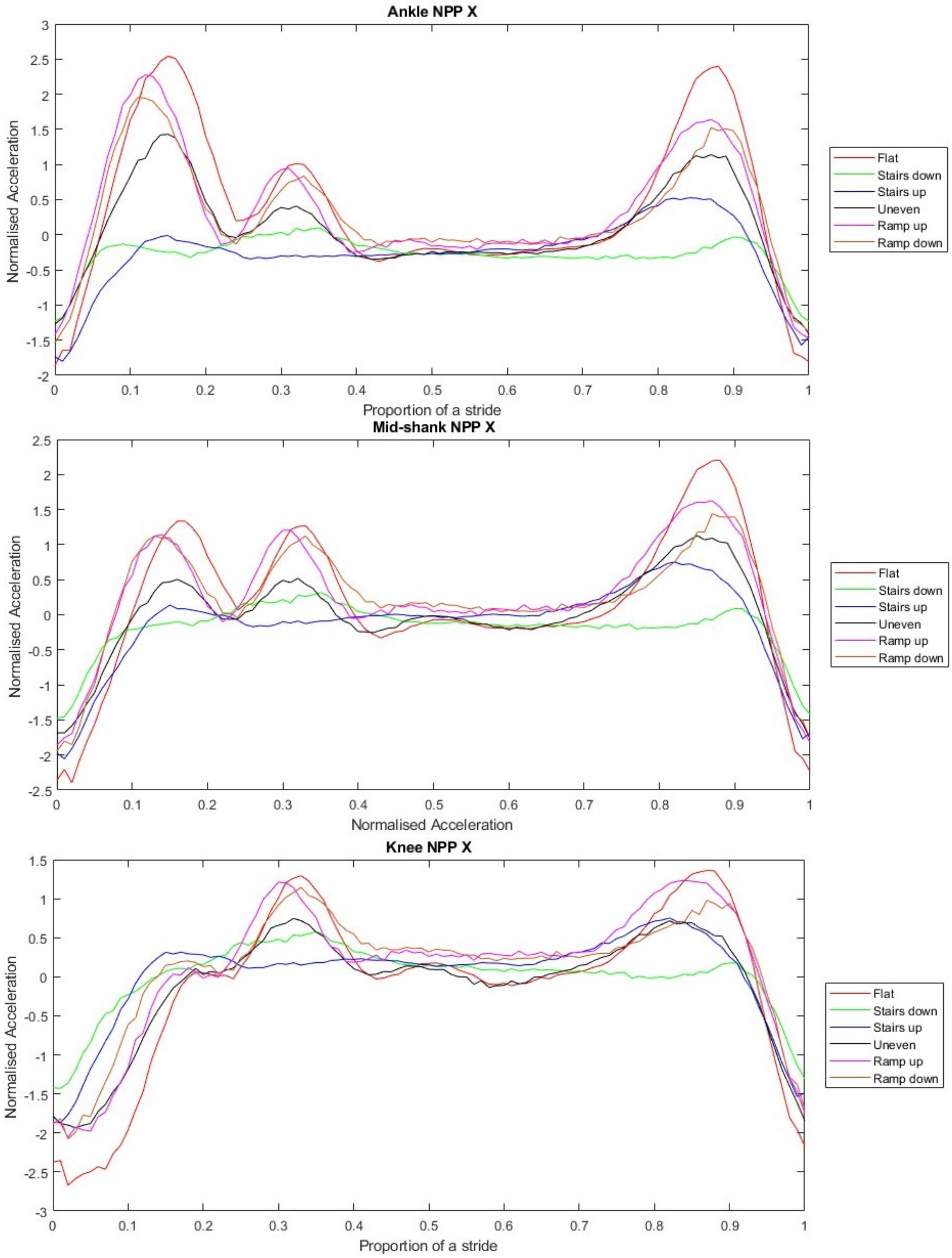

Fig. 5.11 NPP X for ankle (top), mid shank (middle), and knee (bottom). There is variation for stair descending to the other terrains for all placements at about 0.8 of the stride, but the variation is lower for the knee (bottom). The ankle (top) also has clear variation for stair ascending and descending to the other terrains at about 0.15 of a stride.

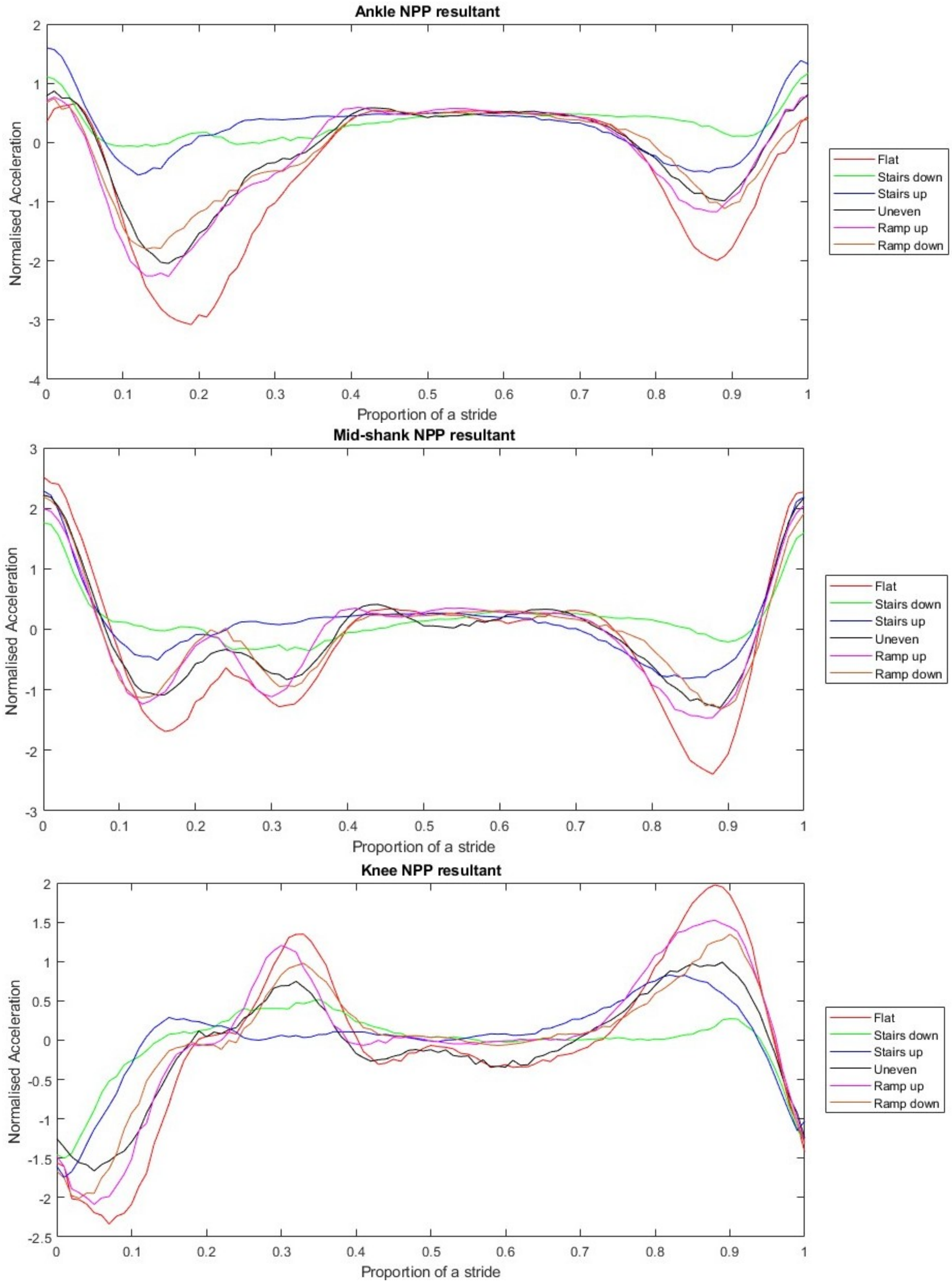

Fig. 5.12 NPP resultant for ankle (top), mid shank (middle), and knee (bottom). There is clear variation between stair descending and the other terrain at about 0.8 of the stride for all locations. There are also variations between flat and the other terrains at this point for the ankle (top) and mid-shank (bottom), with 0.2 of the stride being another place for this variation for the ankle.

In summary, from the plots it can be seen that there is larger variation at the ankle between stair ascending and descending to the other terrains for normalised per person free accelerations in the Z axis, accelerations in the X axis and resultant accelerations. The knee had larger variation for stair ascending and descending for normalised per person free accelerations in the Y axis. The only other terrain that showed variation was flat. There was variation between flat and the other terrains for the ankle and mid-shank in normalised per person resultant accelerations, and at the knee for normalised per person free accelerations in the Z axis. From these plots no location has clear variations for ramp ascending and descending and uneven terrain. The ankle location seemed to show the larger variation for most of the features, but it cannot be said that the ankle will produce higher classification accuracies.

Variance Assessments

A larger variance in IMU acceleration signals across different terrains should produce a high classification accuracy. To investigate this, variance tests were run between terrains for each simulated IMU position. Kruskal-Wallis tests were run on the data to produce p values between each terrain pair for the three IMU placements. The Kruskal-Wallis tests were run for all features with the strides broken down into one hundred segments. The minimum p value, the feature that produced it, the percentage of the stride it occurred and the calculated SMD are presented in Tables 5.9, 5.11 and 5.13, with a detailed breakdown in appendix E.1. The p values were all significant (p<0.0001).

The SMD was also calculated where the lowest p values were found. For the ankle the SMD was above 1 for all terrains against stair use, which was not the case for the mid-shank and knee placements. The mid-shank stair descending against uneven terrain produced a SMD of 0.89 and at the knee stair descending against ramp descending and uneven produced SMD values below 1. For all the placements the lowest SMD values were found for ramp against uneven terrain, ramp ascending against ramp descending, flat against uneven terrain and flat against ramp ascending.

Knee IMU placement had worse values when classifying flat terrain against all other terrains. Ankle placement performed best in seven of the fifteen terrain comparisons, the mid-shank was best in five of the comparisons and the knee placement was best for three. This would suggest that placing the IMU at the ankle could improve the classification of terrain but probably not with a significant difference.

Similar to the Kruskal-Wallis tests, the feature and percentage of the stride that produced the largest SMD was also found for comparison between each terrain individually, these are presented in Tables 5.10, 5.12 and 5.14, with a detailed breakdown in appendix -. The ankle produced the largest SMDs for ten of the comparisons, the mid-shank for three and the knee for two. Twelve of the features and percentage of stride points that produce the largest SMD also produced the smallest P values. Thirteen of the highest SMD were on the same feature as the lowest P values, with ten of these within 3% of stride from each other. The smallest P values and largest SMD was the most aligned for the ankle where 5 were the same and 3 were within 3%. Similar to the Kruskal-Wallis, the ankle had SMD values over 1 for all terrains against stair use; mid-shank also produced this result. The knee did not produce a SMD over 1 for stair descending against ramp descending and uneven terrain. The same terrain comparisons also produced the lowest SMD values as with the p values. At the ankle flat against ramp descending produced a SMD values just below 1 at 0.93.

42% of the most influential features from the Kruskal-Wallis were aspects of the resultant acceleration. Accelerations normalised per person were more influential compared to accelerations that have not been normalised, with 69% of the most influential features being normalised per person. There isn't a significant difference between free accelerations and recorded accelerations, with free accelerations being more influential for 54% of the tests.

From this analysis it can be said that the ankle placement produced larger variations when comparing each terrain separately. This would suggest that for higher classification accuracies measuring the accelerations towards the ankle of the prosthetic shank will produce the highest accuracies. Stair use showed the greatest variation when compared to the other terrains, this suggests that it should be possible to obtain higher classification accuracies for stair use than between the other terrains.

	Uneven	Ramp down	Ramp up	Stairs down	Stairs up
Flat	0.61	0.93	0.52	2.52	2.1
Stairs up	1.05	1.25	1.42	0.79	
Stairs down	1.06	1.55	1.36		
Ramp up	0.55	0.61			
Ramp down	0.5				

Kruskal-Wallis
High
2.5
2
1.5
1
0.5
Low

Table 5.9 SMD for lowest P values when comparing accelerations at the ankle of different terrains

	Uneven	Ramp down	Ramp up	Stairs down	Stairs up
Flat	0.68	0.93	0.72	2.52	2.27
Stairs up	1.05	1.32	1.86	0.79	
Stairs down	1.16	1.58	2.05		
Ramp up	0.65	0.61			
Ramp down	0.52				

SMD
High
2.5
2
1.5
1
0.5
Low

Table 5.10 Highest SMD when comparing accelerations at the ankle of different terrains

	Uneven	Ramp down	Ramp up	Stairs down	Stairs up
Flat	0.66	0.82	0.78	1.66	1.63
Stairs up	0.95	1.12	1.4	0.8	
Stairs down	0.89	1.06	1.57		
Ramp up	0.46	0.46			
Ramp down	0.61				

Kruskal-Wallis
High
2.5
2
1.5
1
0.5
Low

Table 5.11 SMD for lowest P values when comparing accelerations at the mid-shank of different terrains.

	Uneven	Ramp down	Ramp up	Stairs down	Stairs up
Flat	0.66	0.82	0.78	1.76	1.63
Stairs up	1.03	1.25	1.57	0.8	
Stairs down	1.02	1.06	1.76		
Ramp up	0.55	0.53			
Ramp down	0.62				

SMD
High
2.5
2
1.5
1
0.5
Low

Table 5.12 Highest SMD when comparing accelerations at the mid-shank of different terrains.

	Uneven	Ramp down	Ramp up	Stairs down	Stairs up
Flat	0.44	0.61	0.5	1.47	1.46
Stairs up	1.19	1.28	1.84	0.74	
Stairs down	0.67	0.85	1.33		
Ramp up	0.4	0.54			
Ramp down	0.37				

Kruskal-Wallis
High
2.5
2
1.5
1
0.5
Low

Table 5.13 SMD for lowest P values when comparing accelerations at the knee of different terrains.

	Uneven	Ramp down	Ramp up	Stairs down	Stairs up
Flat	0.5	0.63	0.5	1.53	1.53
Stairs up	1.21	1.28	1.96	0.74	
Stairs down	0.86	0.92	1.42		
Ramp up	0.5	0.57			
Ramp down	0.5				

SMD
High
2.5
2
1.5
1
0.5
Low

Table 5.14 Highest SMD when comparing accelerations at the knee of different terrains.

Machine Learning

Up to this point in the analysis, each feature was looked at individually, the machine learning algorithms used in Chapter 4 will also look at interactions between features so the previous analysis can only give an indication on the most influential features and the location that could enhance these features to produce better accuracies. To getter a better understanding on how placement influences accuracies of the terrain classification, the simulated data were used to train and test the machine learning algorithms used in Chapter 4.

The mean accuracies of the 5-fold cross-validation trials for each algorithm and IMU placement are presented in Table 5.15. There is no significant difference between the accuracies for the different placements found from Kruskal–Wallis tests. Ankle placement performed better for RF and SVM compared to Knee and mid-shank, but for KNN, which produced the best accuracies, all placements produced similar results (ankle M=83.59% SD=4.43, mid-shank M=82.91% SD=1.95, knee M=83.25% SD=2.39, $X^2(2)=0.57$ $p=0.75$). This is visualised in Figure 5.13.

Kruskal–Wallis tests were run on the results, as Shapiro-Wilk tests and variance check showed that the data did not meet the requirements for AVNOA, and found significance between the positions for all models (RF $X^2(3)=13.03$ $p=0.0046$, KNN $X^2(3)=11.14$ $p=0.011$, SVM $X^2(3)=13.15$ $p=0.0043$). Dunn's post-hoc tests were run and when combining all the sensors the accuracies reduced significantly compared to ankle and mid-shank for SVM, ankle for RF and knee for KNN. Results of significance test are displayed in Table 5.16. There was no significance between the 3 locations individually, but the ankle placement produced the highest mean accuracy for all models. This shows that consistent location will improve the classification accuracies.

Table 5.15 Terrain classification accuracy results.

Location	RF	KNN	SVM
Ankle	83.08%	83.59%	82.22%
Mid-shank	81.03%	82.91%	81.88%
Knee	79.15%	83.25%	78.80%
Combined	70.43%	70.60%	74.19%

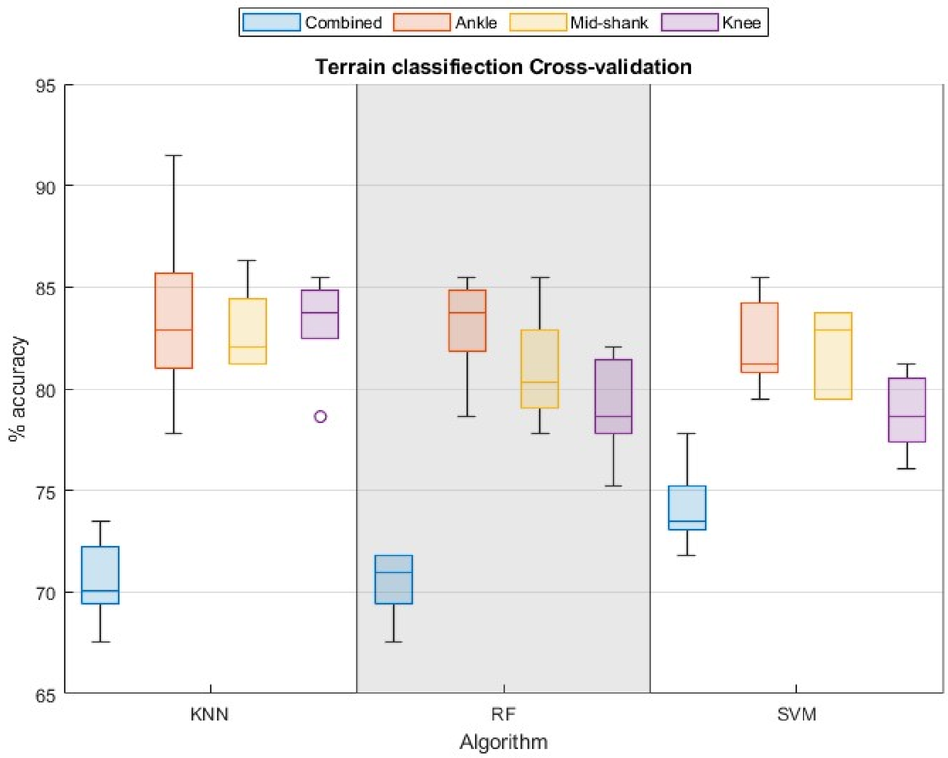

Fig. 5.13 Classification accuracy results. For all models the combined data produced lower accuracies with significant difference to the ankle and mid-shank for SVM, ankle for RF and knee for KNN.

Table 5.16 Significant differences in mean accuracies for Terrain classification for test between location. Mid = Mid-shank

Model		Mean (%) ± SD		Mean (%) ± SD	Difference (%)	*p*-value
RF	All	70.43 ± 1.58	Ankle	83.08 ± 2.38	12.65	0.003
SVM	All	74.19 ± 1.98	Ankle	82.22 ± 2.12	8.03	0.008
	All	74.19 ± 1.98	Mid	81.88 ± 1.98	7.69	0.017
KNN	All	70.60 ± 1.99	Knee	83.25 ± 2.39	12.65	0.016

In summary, the analysis using mRMR, NCA, and RFE identified that the most influential features for terrain classification were largely related to normalized accelerations, but all aspects of the accelerations are important (X, Y, Z, and resultant). This suggested that normalizing the data per participant generally increased the feature's importance, highlighting the benefit of individualized normalization in improving classification accuracy. Plots of mean accelerations show higher variance at the ankle placement. Variance tests also showed higher variances overall at the ankle location when comparing between terrains. It was shown that using data from a consistent position will produce higher classification accuracies. The Ankle position did produce slightly higher mean accuracies but not significantly higher. The subsequent sections will report on the same outcomes but for walking aid use.

5.3.2 Walking aid use.

Feature Importance

The same procedure as for terrain classification was carried out for walking aid use recognition. As LR was also used for walking aid use recognition in Chapter 4, RFE was also performed on LR algorithms which means 30 different variables contributed to the feature importance score. Table 5.17 presents the scored results for feature importance for walking aid use recognition.

Table 5.17 Feature importance for walking aid use recognition. Acceleration feature highlighted in grey. Total score, average score, and median score is displayed for each feature.

Features	Total	Average	Median
Cadence	103	5.72	4
NPP Velocity Y	151	8.39	7
Gyroscope X	154	8.56	7
Velocity Y	156	8.67	6
Velocity X	169	9.39	9.5
NPP Y	184	10.22	12
NPP Velocity Z	188	10.44	9
Z	193	10.72	11
Velocity Z	195	10.83	9.5
Gyroscope Z	200	11.11	11
NPP Gyroscope Y	202	11.22	13
Resultant	207	11.50	10
Y	212	11.78	11.5
NPP Gyroscope Z	217	12.06	12
Gyroscope Y	218	12.11	11
NPP Gyroscope X	218	12.11	12
NPP Cadence	236	13.11	15.5
NPP Z	242	13.44	13.5
NPP resultant	244	13.56	15
X	278	15.44	16
NPP Velocity X	293	16.28	15.5
NPP X	294	16.33	18.5

The delta velocity is the change in velocity between datapoints, acceleration can be calculated by the delta velocity divided by the delta time, if the delta time is constant the delta velocity should be the same but in proportion with the sampling rate. The acceleration output from the IMU is the acceleration recorded at a datapoint. There can be difference between the raw delta velocity and acceleration data, but once filtered and normalised the delta velocity and recorded acceleration should be exactly the same. Figure 5.14 shows the mean and standard deviation for normalised per person acceleration and velocity along the X axis against proportion of a stride, for the data collected in Chapter 4. The plots are basically identical and for the virtual sensors, the sampling rate was always constant at 100Hz, so for this research velocity data was then ignored.

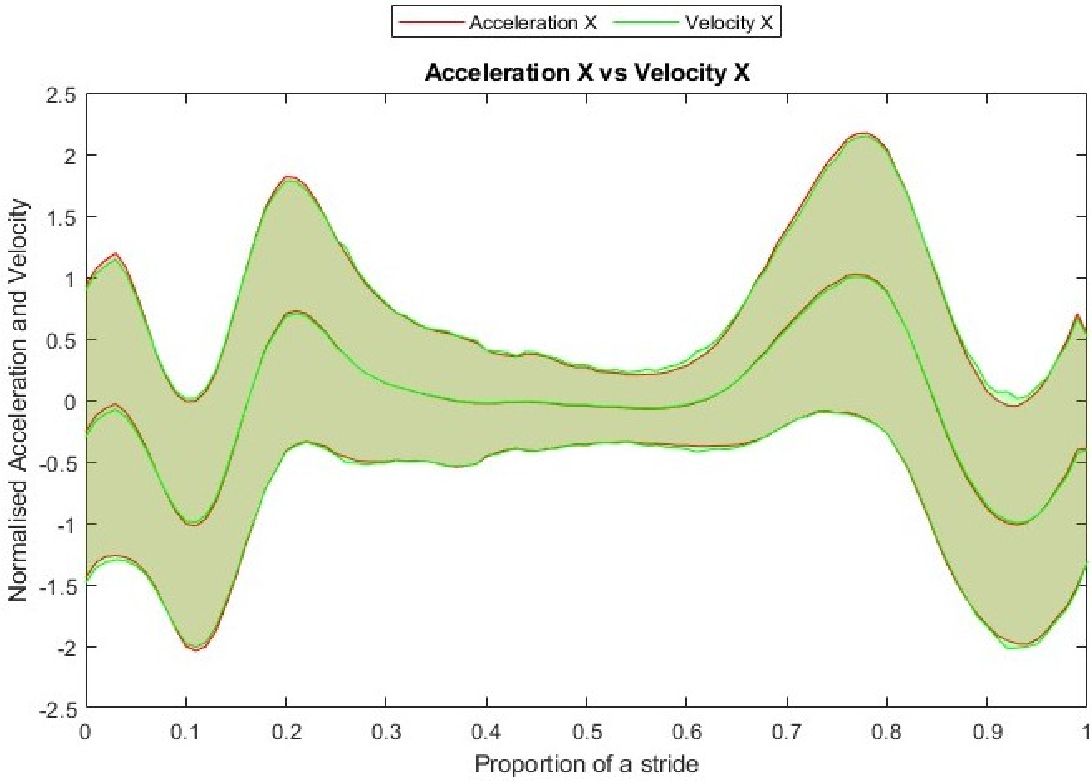

Fig. 5.14 Acceleration X and Velocity X with +/- 1 standard division bands. These features produced identical signals.

Table 22 presents just the acceleration data from Table 5.18.

Table 5.18 Acceleration importance from Table 5.21. Total score, average score, and median score is displayed for each feature.

Features	Total	Average	Median
NPP Y	184	10.22	12
Z	193	10.72	11
Resultant	207	11.50	10
Y	212	11.78	11.5
NPP Z	242	13.44	13.5
NPP resultant	244	13.56	15
X	278	15.44	16
NPP X	294	16.33	18.5

Unlike the terrain classification, the X component of the acceleration does not have a significant influence on the accuracies. All other aspects of the acceleration are suggested to be influential to the accuracy of walking aid use recognition. The normalisation does not seem to have as big an influence as for terrain classification. Plots for the three features suggested to be the most influential are presented in Figures 5.13 to 5.15. The plots show the mean measures against proportion of a stride and a standard difference variation for strides with and without a walking aid.

There does not seem to be a significant variation in Figure 5.15, normalised per person Y axis accelerations, when comparing strides with and without a walking aid, for all of the virtual sensor placements. The largest variation seems to be for the ankle at about 0.95 of the stride.

In Figure 5.16, Z axis accelerations, there does not seem to be a significant variation when comparing strides with and without a walking aid, for any of the virtual sensor placements. The largest variation is from the ankle at 0.2 of the stride.

Figure 5.17, resultant accelerations, does not seem to show significant variation when comparing strides with and without a walking aid, for any of the virtual sensor placements. The largest variation seems to come with the ankle at about 0.15 of a stride.

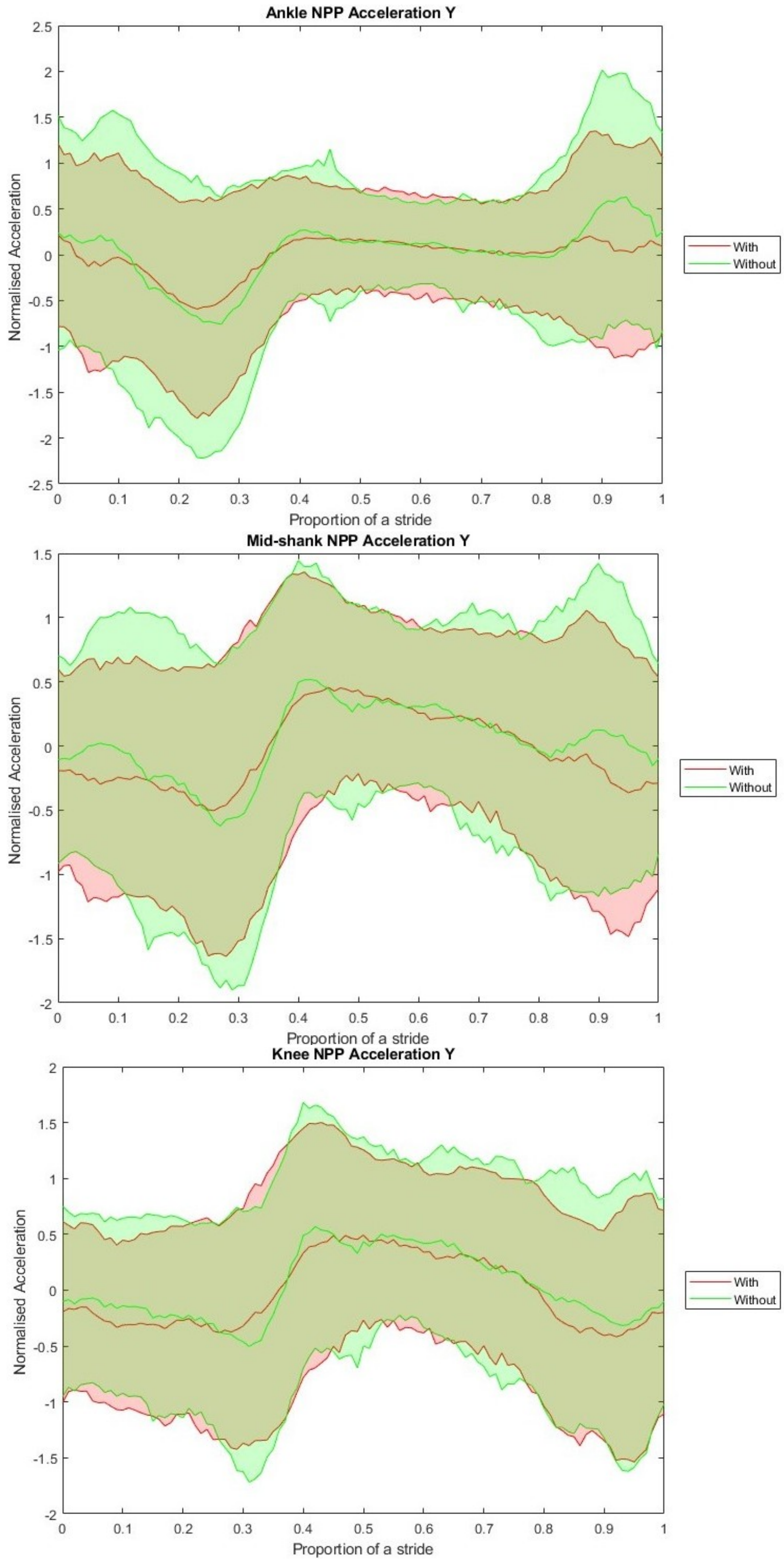

Fig. 5.15 NPP acceleration Y for ankle (top), mid shank (middle), and knee (bottom) with +/- 1 standard division bands. There does not seem to be a significant variation in the normalised per person Y axis accelerations when comparing strides with and without a walking aid, for any of the virtual sensor placements. The largest variation seems to be for the ankle at about 0.95 of the stride.

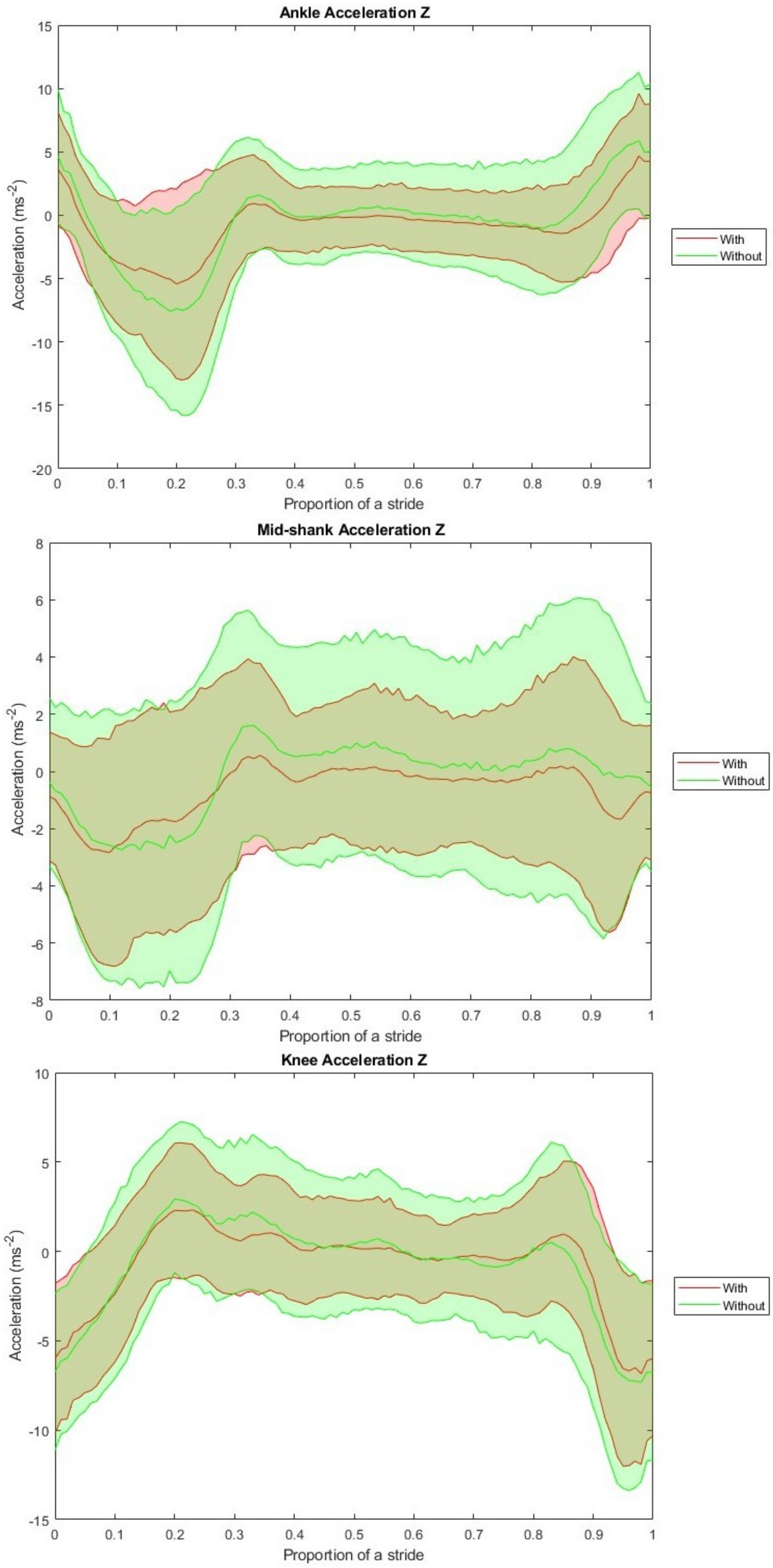

Fig. 5.16 Acceleration Z for ankle (top), mid shank (middle), and knee (bottom) with +/- 1 standard division bands. There does not seem to be a significant variation in the Z axis accelerations when comparing strides with and without a walking aid, for any of the virtual sensor placements. The largest variation is from the ankle at 0.2 of the stride.

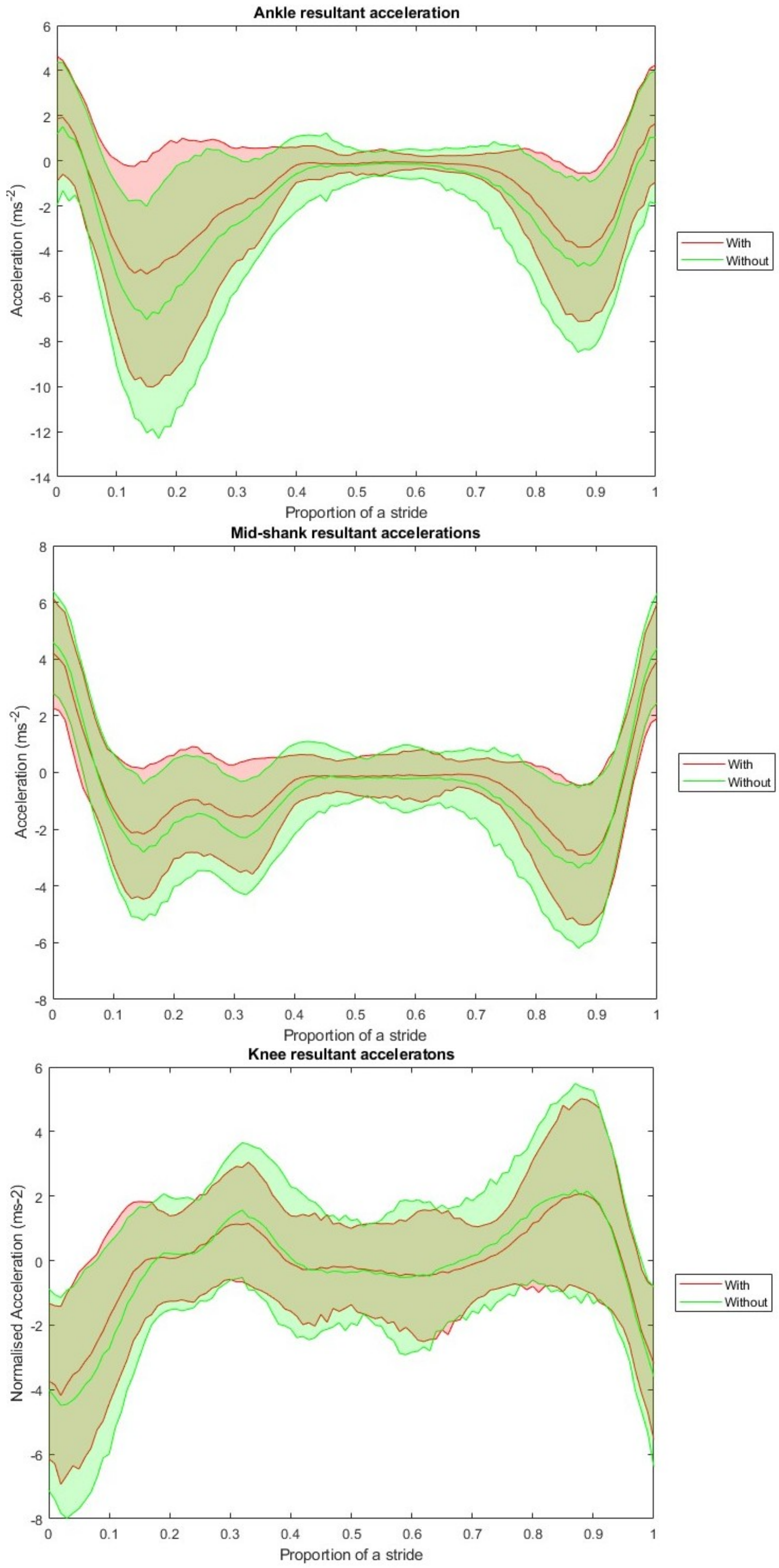

Fig. 5.17 Resultant accelerations for ankle (top), mid shank (middle), and knee (bottom) with +/- 1 standard division bands. There does not seem to be a significant variation in the resultant accelerations when comparing strides with and without a walking aid, for any of the virtual sensor placements. The largest variation seems to come with the ankle at about 0.15 of a stride.

In summary the plots do not show a significant difference between strides with and without a walking aid. The ankle seemed to produce larger variation between walking aid use and none use but this still does not appear significant.

Variation Assessments

A Kruskal-Wallis test was carried out in the same way as for the terrain classification, section 5.3.1, and the results are presented in Table 5.19. As with the terrain classification, the lowest p values for each location were significant (p<0.0001). The feature and percentage of a stride that produced the largest SMD for all three locations was the same as the feature and percentage of a stride that produced the lowest P value.

The highest SMD for all locations were small and did not suggest a significant difference. For the knee and mid-shank location, the largest SMD was for the resultant accelerations at 1% of the stride;, looking at Figures 5.17 there does not seem to be a large variation at this point which would align with the low SMD. For the ankle, the point of highest SMD was for the Z axis accelerations at 94% of the stride, looking at Figure 5.16 there does seem to be slight variation at this point but not significant which aligns with the low SMD. These results would suggest that ankle accelerations would produce better walking aid recognition accuracies but that accelerations alone will not produce high classification accuracies.

Table 5.19 Results for Kruskal-Wallis and SMD tests. All SMD are small which suggests low variance between strides with and without a walking aid.

Position	Ankle	Mid-shank	Knee
Percentage of stride	94%	1%	1%
Feature	Z	Resultant	Resultant
SMD	0.284	0.251	0.237

Machine Learning

As with the terrain classification, section 5.3.1, the simulated accelerations were used to train and test the machine learning algorithms used in Chapter 4. Table 5.20 display the mean accuracies for the 5-fold cross-validation for the 3 simulated locations and trials run when the data of the 3 locations are combined and Figure 5.18 visualises these results. The accuracies would suggest that a knee-placed IMU would be more accurate at recognising walking aid use when using a RF or SVM algorithm. The mean accuracy for SVM for the knee placement is the highest. Combining the three locations seems to produce the worst accuracies. Kruskal–Wallis tests were run on the results, as Shapiro-Wilk tests and variance check showed that the data did not meet the requirements for AVNOA, and found significance between the positions for SVM models (RF $X^2(3)=5.67$ $p=0.13$, SVM $X^2(3)=8.74$ $p=0.033$). Dunn's post-hoc tests were run and the only significant result was that the knee accelerations are significantly better than the combined accelerations for SVM (Knee M=76.07% SD=6.33, Combined M=67.01% SD=2.79, difference=9.06% $p=0.026$).

Table 5.20 Walking aid classification accuracy results.

Location	RF	SVM
Ankle	69.91%	70.26%
Mid-shank	70.26%	71.79%
Knee	74.19%	76.07%
Combined	65.47%	67.01%

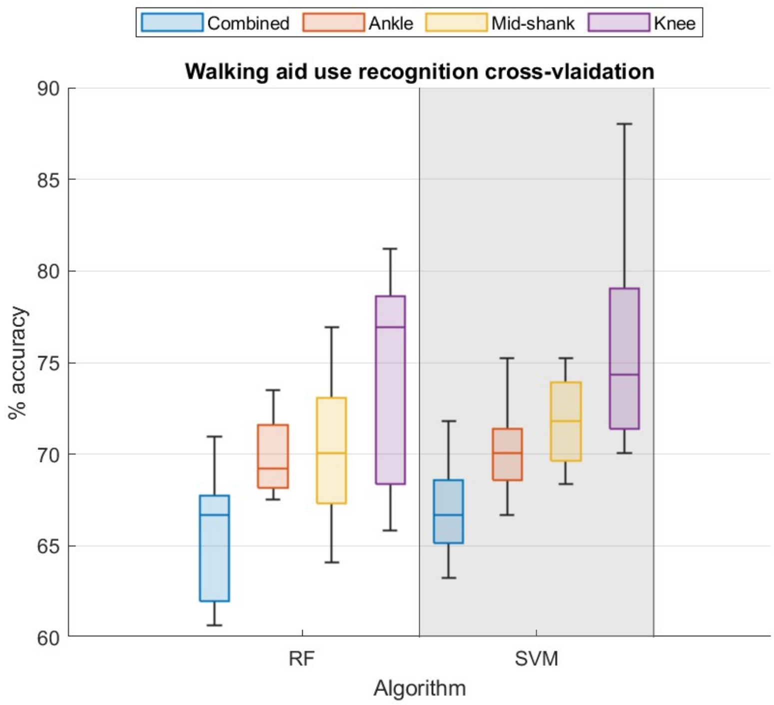

Fig. 5.18 Classification accuracy results. Combined data produce the lowest accuracies and significantly lower than the knee for SVM.

5.4 Discussion and Conclusions

5.4.1 Terrain classification

The results for the terrain classification analysis did not show conclusively where the ideal location would be in order to collect acceleration data from a prosthetic shank that produces the highest classification accuracies. It can, however, be said that a consistent location would improve the accuracies. This backs up evidence presented by Ruder [183], Tan [184] and Lutzner [185] that found small deviation in IMU placement can cause discrepancies in the accuracies produced.

Stair use produced the greatest variability when compared to the other terrains, which would suggest that stair use would produce higher classification accuracies. This was evident when looking at the confusion matrices. Figure 5.19 displays the confusion matrix for the KNN classifier using the ankle accelerations. Class 2 is stair ascending and class 3 is stair descending. Stair use produced the highest classification accuracies. Ramp ascending (class 4) produced the lowest classification accuracies and was misclassified as flat, the most common, and had uneven misclassified as ramp ascending frequently. These results match the SMD results from section 5.3.1. These insights gained are in agreement with Jamieson [40].

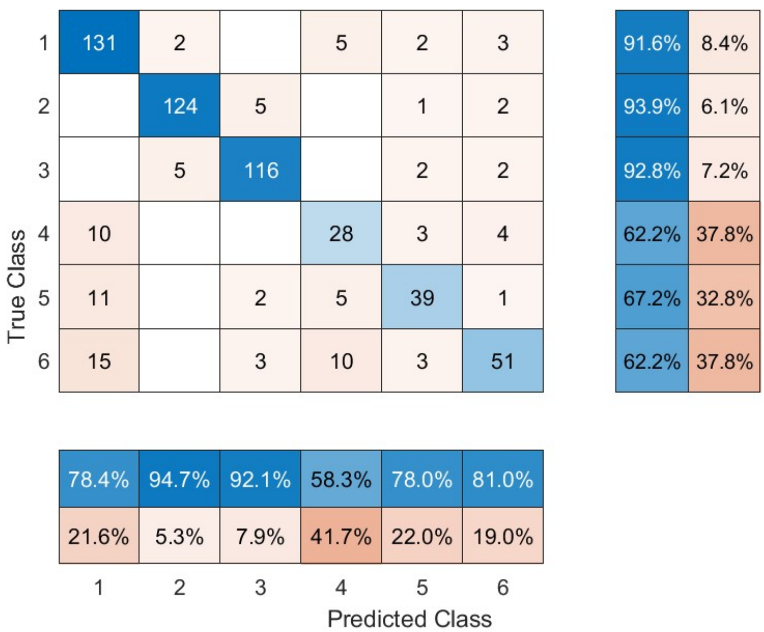

Fig. 5.19 Confusion matrix Ankle KNN, 1=flat, 2=stairs ascending, 3=stairs descending, 4=ramp ascending, 5=ramp descending, 6=uneven. Stair use (class 2 and 3) produce the highest classification accuracies and ramp ascend (class 4) produces the lowest.

5.4.2 Walking aid use

Like for the terrain classification, the analysis did not conclusively find the ideal location for acceleration to be measured on a prosthetic shank to produce the highest walking aid use recognition accuracies. The knee acceleration produced the highest classification accuracies, but the ankle seemed to show the most variance in the data. Also mirroring the terrain classification, having data from varying locations had a negative effect on the accuracies. All the confusion matrixes showed similar results where the classification models did not overclassify, and produced similar accuracies for strides with and without a walking aid. This is demonstrated in the knee SVM confusion matrix, Figure 5.20, where the split is even between the classes. This is contrary to the results shown in Chapter 4 where the models overclassified to walking without a walking aid. The difference in the dataset for this study was that the split between strides with and without a walking aid was even (293 without a walking aid and 292 with a walking aid).

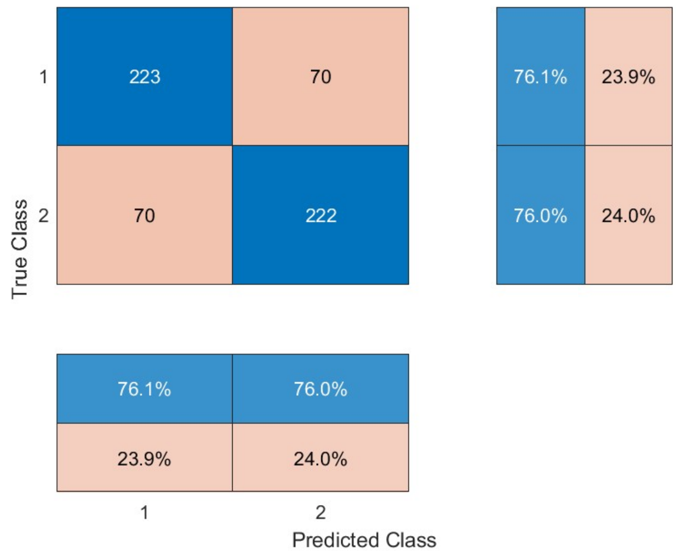

Fig. 5.20 Confusion matrix SVM Knee, 1=without walking aid, 2=with walking aid. The accuracies are consistent for both walking aid use (class 2) and non-use (class 1), which indicates that the model is not overfitting.

Notably, the accuracies produced in this study and Chapter 4 for walking aid recognition appear not acceptable for clinical use. The low levels of variance in the acceleration data are a contributing factor to this, and we question whether it would be possible to accurately classify walking aid use with just an IMU.

5.5 Limitations

Due to the restrictions in the motion capture laboratory, this research could only investigate flat ground, uneven ground, ramp and stair traversing. Chapter 4 also looked at grass and unstable (gravel) terrain. This means that the results of this study will not completely align to the models that were produced in Chapter 4 and how the untested terrains affect the accuracies of the tested terrains remains unknown. From the results in Chapter 4 there were only a small number of strides that were misclassified between stair and ramp use to unstable and grass. There was some misclassification between flat and uneven terrain strides to grass. Not many of the strides on one of the tested terrains were misclassified as unstable but unstable was misclassified as flat 39% of the time. Taking this into consideration, it can be assumed that the untested terrains would not have a huge effect on the accuracies of the tested terrains.

Only strides where the participant had toe off and the next heel strike on the terrain were counted for this research. Therefore, transition strides were not included. This was to eliminate these as a potential source of error in the findings. Due to restriction in the motion capture laboratory only a short section of uneven terrain and ramp could be used and meant that only one or two strides would be recorded for these terrains per trial. This is evident when looking at the number of strides per terrain, Table 5.21. This could be affecting the accuracies in this study for these terrains.

Table 5.21 Strides recorded per terrain

Terrain	Number of strides
Flat	143
Stair ascending	132
Stair descending	125
Ramp ascending	45
Ramp descending	58
Uneven	82

This research was carried out in a motion capture laboratory under idealised conditions. These results may hence not necessarily be replicated in real-world conditions.

In conclusion:

Although no conclusive location to place a prosthetic shank-mounted IMU was determined for producing higher terrain classifications or walking aid recognition accuracies, the significant impact that sensor placement consistency has on classification accuracy has been demonstrated. The findings provide valuable insights into how terrain and walking aid usage can be distinguished, especially with the ankle placement showing promising results in terrain classification, particularly for stair use. Unfortunately, the accuracies produced in this study and Chapter 4 for walking aid recognition seem too low to be useful for clinical decision-making. The low levels of variance between walking aid use found in this study could explain the low accuracies.

Nevertheless, this chapter's attempt to refine the system and improve classification accuracies produced useful insights.

The next chapter (Chapter 6) will conclude this body of work with a clinical evaluation of the system.

Chapter 6

Clinical evaluation of prosthesis users' real-world activity data recorded from the objective sensor system.

6.1 Background

In the previous two chapters (Chapters 4 and 5), a sensor system and associated algorithms were designed and refined, and associated accuracies reported. The **purpose** of this chapter was to assess if real-world data collected with the objective sensor system for lower-limb prosthesis users meets clinical needs - specifically, if the system's outputs align with the requirements identified by clinicians in Chapter 2.

Based on the sensor classification models developed in Chapter 4 and the initial clinician input in Chapter 2, this chapter uses real-world longitudinal data to assess the system's clinical utility. In Chapter 2, interviews conducted with clinical experts were analysed and outcomes influenced the direction of the research that was subsequently presented in Chapters 3, 4 and 5. The interviews set out the needs and requirements that this research aimed to fulfil. To assess whether those requirements are met by the system, longitudinal real-world data, collected from lower limb prosthesis users, and associated output on free-living activities needed to be shared with clinicians together with standard clinical outcomes. A final set of interviews with the clinicians was able to informed whether the research delivered a system that met their clinical needs.

In support of this approach, a previous research study started by conducting qualitative research with clinical experts before development of an activity tracking system [40], however in this study, no qualitative research after the system was developed took place. The authors acknowledged that this was a drawback to their research and development. To the author's knowledge no comparable research of a similar nature to this project exists that conducted a final qualitative assessment to evaluate the technology developed. Yet, qualitative data collection and analysis have been shown to be a powerful tool for technology development in healthcare [195]. Conducting qualitative research to not just influence design and development but to evaluate and refine products, and to ensure they meet the needs of a target user is vital to ensure uptake. The target users for the system this research has developed are clinicians that conduct K level assessments and who prescribe prosthetic components. Hence obtaining their views regarding the system and the data it produces is paramount to verify that the system is indeed meeting their needs, and to inform on any further development.

6.1.1 Aims of Chapter 6

The overarching aim of this chapter was to determine whether the system's real-world data outputs meet the practical needs and requirements of clinicians working with lower-limb prosthetic users.

The specific aims were:

- To use the new sensor system to collect longitudinal real-world activity data for three amputees, share the outcomes in the form of short activity reports with clinicians.
- To conduct semi-structured interviews with clinicians to obtain their views on the sensor system and its outputs.

6.2 Methods

To investigate whether the IMU system developed in Chapter 4 has clinical relevance, semi-structured interviews were conducted with four clinical experts. To be able to demonstrate the capabilities of the system, three lower limb prosthetic users were first monitored with the system over a two-week period. The prosthetic users' data were then classified using the classification models developed in Chapter 4. To investigate whether the system data may provide additional insights to those obtained through standard clinical assessments, each prosthetic user was also assessed by a clinician using a standard K level assessment. The interviewees were then presented with a summary of the clinical assessment and a short report that summarized the system output for each participant at least 24 hours before the interview took place.

Ethical approval was granted to collect the real-world activity data and conduct the interviews (Ethical approval numbers for University Ethics: 4743, Appendix D.2).

6.2.1 Data collection

Participants

To ensure that the classification accuracies were as high as possible. it was decided to only recruit participants that had participated in the earlier data collection that had served to create the classification models, as discussed in Chapter 4. This reduced the variability of the data compared to the data that trained the models. Out of the twenty lower limb prosthetic users who participated in the data collection, three were able to participate in this longitudinal real-world monitoring study and provided informed consent. Details of the participants are given in Table 6.1. All three participants were male and above 60 years old. Two were transtibial and one was transfemoral.

Table 6.1 Participant details.

Participant	Participant number from Chapter 4	Age	Sex	Prosthetic type
1	2	64	Male	Transtibial
2	1	86	Male	Transtibial
3	10	60	Male	Transfemoral

Sensors

The Xsens Awinda used for the data collection that created the classification models cannot collect longitudinal data, so this type of IMU was not suitable for this study. Instead, x-io IMU 3s were used, because they can store data for up to 20 days and have a battery life of 13 hours, whereas the Xsens Awinda cannot store data and only has a 6-hour battery life. The 13-hour battery life was the largest found for IMUs that matched the Xsens Awinda, but it did mean that during the data collection the participants had to charge the IMU at night while they were not using their prosthetic. The x-io IMU3 can record at a sampling rate up to 400Hz, but as found in Chapter 4 only a sampling rate of 20Hz had to be used given the frequency content of prosthetic gait kinematics. Table 6.2 displays the specification comparison between the x-io IMU 3 and the Xsens Awinds.

Table 6.2 IMU specification comparison data.

	x-io IMU3	Xsens Awinda
Gyroscope range	±2000°/s	±2000°/s
Gyroscope noise	$0.014°/s/\sqrt{Hz}$	$0.01°/s/\sqrt{Hz}$
Accelerometer range	±235m/s2	±160m/s2
Accelerometer noise	190 µg/$\sqrt{Hz}$	200 µg/$\sqrt{Hz}$
Magnetometer range	±1.3 Gauss (X,Y) ±2.5 Gauss (Z)	±1.9 Gauss
Magnetometer noise	0.3 mGauss/$\sqrt{Hz}$	0.2 mGauss/$\sqrt{Hz}$
Static accuracy (roll/pitch)	0.5° RMS	0.5° RMS
Static accuracy (heading)	1° RMS	1° RMS

Real-world data collection with sensor system

For the data collection the participants attended an individual training session at the University of Salford. During this session the participants were taught how to charge the IMU, how to remove and attach the IMU to their prosthesis shank and what the IMU warning lights meant. A housing was designed to attach the IMU to a prosthesis shank that ensured the orientation of the IMU match the orientation used in the data collection documented in Chapter 4. The 2-week data collection started at the end of the training session for each participant. At the end of the data collection the participants were given the choice to either post the IMU to the University of Salford via a prepaid envelope provided to them, or return the IMU when they came in for their clinical assessment.

Clinician Interviews

The aim of the semi-structured interviews was to gather an understanding of the opinions of clinical experts on the objective system and its output that had been developed to assist them with lower limb prosthetic prescriptions. This involved obtaining the interviewees views on the data that the system can produce, its clinical relevance, if it could affect a lower limb prosthetic prescription, how the data is being displayed and any improvement or changes that could be made to the system. The rationale for using semi-structured interviews was the same as for the initial interviews conducted in Chapter 2. It was important to obtain individual views and not just a general consensus, and with the small number of eligible clinicians, conducting the research through interviews made recruiting easier as the researcher could meet individuals when it suited them.

The development of the interview guide was influenced by the work of Kallio et al. [86], to obtain the necessary information and to give the participants an understanding of the questions that would be asked and why. The interview guide (Appendix C.2) was sent to all participants before the interviews to give participants an understanding of the type of questions that would be asked and why.

All interviews were conducted via video call and audio was recorded using an external General Data Protection Regulation (GDPR) compliant recording device. The audio was then transcribed, and any identifying details of the participants were removed from the transcripts. The recordings were transcribed using speech to text software and then manually checked and corrected where needed. The transcripts were analysed using thematic analysis based on a framework approach described by Braun and Clarke [87].

6.2.2 Data analysis

Activity Periods

As data were recorded continuously, there were sections of the collected data where the participant was not walking using their prosthesis. As the classification models have only been trained to classify stride data, the stride data had to be extracted from the collected data. To do this, first the strides were identified, in the same way as in Chapter 4, by identifying peaks in the Y component of the gyroscope data. To ensure that only stepping periods were included, data had to have 3 peaks in a 7.5s period. This eliminated peaks that were not strides. To align with most common activity monitors, the stride count was transferred to step count through multiplying by 2.

Activity classification

For terrain classification, as stated in Chapter 4, the classification model that was used was an optimised RF model, and for the walking aid recognition the optimised SVM model was used. The data were split into windows that contained a whole stride, each of which contained 10 datapoints.

The x-io IMU3 units do not output delta velocity but as discovered in Chapter 5 section 5.3.2 the delta velocity data were similar to the acceleration data in any case. However, to check whether there were any implications of using IMUs that do not output delta velocity, the data collected in Chapter 4 were re-run with the SVM model that was then to be used in this study without the delta velocity data. Reassuringly, the mean accuracies for walking aid recognition for both leave-some-out and cross-validation were similar to the results with delta velocity from Chapter 4, see Table 6.3. All other features used matched the optimised models set out in Chapter 4 section 4.3.2 for terrain classification and section 4.3.3 for walking aid recognition.

Table 6.3 SVM walking aid recognition percentage accuracy comparing IMU input data with and without delta velocity output.

	Leave-some-out	Cross-validation
With Velocity	65.03%	84.42%
Without Velocity	65.23%	83.85%

Standard K level assignment

One clinician conducted all three clinical assessments and assigned K levels to the 3 prosthesis users as per standard practice. The assessments were conducted at the University of Salford in the prosthetic and orthotic clinic room, to be as similar as possible to current clinical practice. The clinician had 8 years of experience conducting K level assessments for prosthetic component prescription. The assessments were twenty minutes long, as this was the same length as the clinician would usually have to conduct clinical assessments. The assessments were conducted as close to a clinical assessment as possible, and consisted of the clinician having a conversation with the prosthetic user about their activity and medical history, and ended with the clinician deciding on an appropriate K level for each participant.

Lower limb prosthetic participant report

For each participant, a report was compiled. The report included a summary of the clinical assessment that included the information that the clinician deemed key for deciding the participants K level. The report also included a summary of the participants activity data, which was displayed in 2 parts. The first part was a summary of the data that included the average daily steps in total and on each terrain, the average recorded cadence and a percentage of the steps that were at a cadence 20% higher than the average, and the percentage of steps that used a walking aid. The amount of steps at a cadence 20% higher than mean cadence was included as the ability to vary cadence was highlighted by the clinical experts in Chapter 2 and the ability to increase cadence requires more energy consumption and demonstrates a higher activity level [196]. The second part was a table that displayed data for each day of the data collection period, this included the number of steps on each terrain and in total, the average daily cadence and the percentage of steps that day that were at a cadence 20% higher than the participants' average cadence, and the number of steps taken with and without a walking aid plus the percentage of steps taken with a walking aid. The reports are presented in section 6.3.1.

6.2.3 Clinician Participants

Participants for the interviews were recruited in multiple ways. First the participants of the interviews in Chapter 2 were invited. The interviews in Chapter 2 set the requirements for the system and to see if those requirements were met it was the intention to interview the same clinical experts where possible. Only three of the six from the initial interviews could participate. New clinical experts were therefore then recruited who had not participated in the initial interviews but had to meet the same inclusion criteria. This did, however, have the benefit that three clinicians were able to assess the system output in relation to what they had initially stated is their need, whilst the new clinician was able to provide a fresh, unbiased perspective as would be the case if the system went into clinical practice.

The inclusion criteria were: experience with activity level assessments in users of lower limb prostheses, able to provide informed consent, and able to do an interview over the phone in the English language.

Participants were recruited through the International Society of Prosthetics and Orthotics and links connected to the Centre of Doctoral Training in Prosthetics and Orthotics at the University of Salford. Identified clinicians were emailed the studies participant information sheet and if they were interested in participating to respond to the researcher to arrange the interview. Details of the participants are displayed in Table 6.4.

Table 6.4 Clinician participant information.

Participant number	Participant number from chapter 2	Background	Type of clinic
C1	P4	Amputee specialist physiotherapist - 20 years' experience	NHS
C2	P1	Prosthetist - 27 years' experience	Blatchford
C3	P3	Prosthetist - 16 years' experience	Blatchford
C4	N/A	Prosthetist – 23 years' experience	NHS

6.3 Results

This section presents the results from the data collected with the sensor system on lower limb prosthetic users, as well as clinician feedback on that data obtained from the interviews. The results are organised into two main sections, summaries of the activity data collected from lower limb prosthetic users in the real-world that were provided to the clinicians, and thematic analysis of clinician feedback on the system's clinical relevance and areas for improvement.

6.3.1 Prosthesis user activity data summaries as presented to clinicians

In the next few sections, summaries of the clinical K level assessment, with an assigned K level, as well as a summary from the IMU data output and a breakdown of daily activities are shown for the three prosthesis user participants as they had been presented to the clinician interviewees.

Participant 1

Clinical assessment summary

- Transtibial Left leg
- Cause: Trauma
- 64 years old
- Hobbies include walking (approximately 4 to 5 miles a day)
- Carer for his wife
- Does not play sport
- Does not smoke
- Drinks a bit
- Type 2 diabetes
- Skin condition is good
- Oxford scale 5 for muscle strength in amputated limb
- No knee instability
- Walks unaided
- Is able to don and doff their prosthetic unaided

Assessed as a K3

Summary of recorded data

On average: **7905 steps per day**

Average steps on **grass** a day: **672**

Average **stairs climbed** a day: **254**

Average **stairs descended** a day: **132**

Average steps **up a slope** a day: **614**

Average steps **down a slope** a day: **976**

Average steps on **uneven terrain** a day: **205**

Average steps on **unstable terrain** a day: **244**

Average cadence: **73.9 Steps per minute**

% of steps **above 20% average Cadence**: **27.4%**

% of walking aid use: **0%**

	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Day 8	Day 9	Day 10	Day 11	Day 12	Day 13	Day 14	Day 15
Flat	512	2418	5242	6788	6854	2926	7174	2682	5822	4392	6238	4636	5886	6440	4112
Grass	74	282	424	474	1232	706	978	324	1690	344	1156	786	516	592	496
Up Stairs	92	506	254	370	168	160	174	190	244	308	336	210	214	396	174
Down Stairs	8	32	172	240	164	42	66	60	152	266	174	146	234	112	120
Up Slope	2426	426	62	106	212	1352	1162	172	788	62	1074	298	84	476	512
Down Slope	470	3868	2520	1012	1020	818	1436	464	910	194	340	362	336	560	336
Uneven	10	48	210	308	668	176	522	22	252	82	74	220	162	130	194
Unstable	106	374	492	100	170	178	344	78	280	126	468	152	174	324	288
Total	**3698**	**7954**	**9376**	**9398**	**10488**	**6358**	**11856**	**3992**	**10138**	**5774**	**9860**	**6810**	**7606**	**9030**	**6232**
Cadence steps per minute	65.1	73.9	72.3	73.6	74.4	75.6	79	74.7	77	67.3	72.1	71.5	70.3	74.2	72.3
% above 20% normal Cadence	28.2	3.2	25.6	25.2	29.7	33.4	40.4	35.7	32.4	21.8	31.3	23.8	21.3	30.6	21.5
Without a walking aid	1849	3976	4688	4699	5244	3179	5928	1996	5069	2883	4930	3404	3803	4515	3116
With a walking aid	0	1	0	0	0	0	0	0	0	4	0	1	0	0	0
% walking aid use	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%

Participant 2

Clinical assessment summary

- Transtibial left leg
- Cause: Trauma
- 86 years old
- Has a pacemaker
- Has high blood pressure
- Has asthma
- Has type 2 diabetes
- Does not smoke
- Drinks a bit
- Walks dog daily (approximately 3 to 4 miles)
- Hip replacement in contralateral leg
- No pain in contralateral leg
- Fibular amputated longer than tibia, which does cause reddening of skin at end of residual limb
- Reddening of skin also on tibia head
- Walks unaided
- Is able to don and doff their prosthetic unaided

Assessed to be a K3, but noted that current prescription is a K2 leg

Summary of recorded data

On average: **6562 steps per day**

Average steps on **grass** a day: **967**

Average **stairs climbed** a day: **133**

Average **stairs descended** a day: **66**

Average steps **up a slope** a day: **585**

Average steps **down a slope** a day: **960**

Average steps on **uneven terrain** a day: **419**

Average steps on **unstable terrain** a day: **166**

Average cadence: **91.8 Steps per minute**

% of steps **above 20% average Cadence**: **16.5%**

% of walking aid use: **0%**

	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Day 8	Day 9	Day 10	Day 11	Day 12	Day 13	Day 14	Day 15
Flat	3604	3088	3614	2758	2412	4066	3222	2582	3654	2588	4258	2632	3314	2216	4954
Grass	396	850	296	1288	1610	660	538	954	1544	1228	436	1874	1084	856	886
Up Stairs	132	106	202	126	150	120	90	122	204	76	170	128	136	80	158
Down Stairs	72	62	126	72	66	30	52	60	74	60	102	50	50	28	86
Up Slope	604	516	286	730	632	866	820	624	642	552	290	418	652	652	498
Down Slope	332	712	216	1150	1108	956	966	822	1136	1666	1158	1162	1416	1170	436
Uneven	260	462	102	364	478	264	380	250	1124	470	266	788	506	288	290
Unstable	142	184	136	112	218	102	214	118	270	146	222	152	188	68	220
Total	**5542**	**5980**	**4978**	**6600**	**6674**	**7064**	**6282**	**5532**	**8648**	**6786**	**6902**	**7204**	**7346**	**5358**	**7528**
Cadence steps per minute	95.2	85.8	84.5	87.4	87	94.5	91.1	91.8	96.2	96.4	90.4	93.3	92.6	98.5	90.5
% above 20% normal Cadence	18.4	6.1	0.8	5.4	7.6	16.6	8.7	20.2	25.2	22.2	30.7	17.4	11.4	30.4	21.5
Without a walking aid	2771	2990	2489	3300	3337	3532	3141	2766	4324	3393	3451	3602	3673	2679	3764
With a walking aid	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
% walking aid use	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%

Participant 3

Clinical assessment summary

- Transfemoral left leg
- Cause: Compartment syndrome 3 years after a fall
- 60 years old
- Has type 2 diabetes
- Does not smoke
- Drinks a little
- Lives alone
- Can walk 11Km, regularly walks 5Km
- Does not run and no desire to
- Cycles 12 to 15 miles a day
- Goes to the gym regularly
- Coaches archery
- Desire to play golf

Assessed to be a K4

During assessment had a fall on day 3 that restricted normal activities.

Summary of recorded data

On average: **4032 steps per day**

Average steps on **grass** a day: **208**

Average **stairs climbed** a day: **131**

Average **stairs descended** a day: **12**

Average steps **up a slope** a day: **482**

Average steps **down a slope** a day: **364**

Average steps on **uneven terrain** a day: **35**

Average steps on **unstable terrain** a day: **23**

Average cadence: **73.4 Steps per minute**

% of steps **above 20% average Cadence**: **23.8%**

% of walking aid use: **1%**

	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Day 8	Day 9	Day 10	Day 11	Day 12	Day 13	Day 14
Flat	3492	6692	1770	1944	1500	1684	2462	1252	1402	946	4174	5592	2680	3302
Grass	194	234	142	256	4	164	162	92	82	118	322	344	354	438
Up Stairs	36	1236	50	56	78	60	48	52	38	14	42	36	58	30
Down Stairs	10	50	18	14	4	12	8	12	20	8	4	4	6	4
Up Slope	522	1142	162	208	58	206	212	46	132	78	1064	1578	438	902
Down Slope	518	1088	120	270	24	180	122	42	152	86	510	1292	136	550
Uneven	64	22	18	98	8	16	20	8	24	12	46	64	68	22
Unstable	30	46	18	28	34	20	16	28	10	8	32	18	26	6
Total	**4866**	**10510**	**2298**	**2874**	**1710**	**2342**	**3050**	**1532**	**1860**	**1270**	**6194**	**8928**	**3766**	**5254**
Cadence steps per minute	75.9	77.9	59.1	66.8	65.1	60.3	59.7	54.4	67.7	61.7	78.8	83.5	67.9	80.4
% above 20% normal Cadence	19.6	28.7	8.9	11.8	17.3	9.7	5.0	3.8	24.3	1.9	3.6	54.5	12.6	40.9
Without a walking aid	2432	5247	1129	1428	838	1151	1500	733	924	629	3095	4464	1874	2626
With a walking aid	1	8	20	9	17	20	25	33	6	6	2	0	9	1
% walking aid use	0%	0%	2%	1%	2%	2%	2%	4%	1%	1%	0%	0%	1%	0%

6.3.2 Interview Results

This section presents the results of the clinician interviews, which were conducted to evaluate the real-world activity data of the lower limb prosthetic users that was collected by the sensor system. Thematic analysis of the interview transcripts revealed the following six themes: "General feedback" and "Clinical use", "Presentation of system output", "Comparison of system data with clinical assessments", "Charging", "Areas for improvement", with some of these containing a number of subthemes.

Theme 1: General feedback on the system

This theme had two subthemes, namely "Positives" and "Output data of the system" which are discussed in the following two sections.

Positives

When asked about their general thoughts on the data that the system produced, the interviewees gave very positive responses:

> *"I found the data very helpful."* (C1)

> *"I thought it was really good, honestly, really, the thing is, good stuff."* (C2)

> *"I think it's amazing data. It's really, really impressive. Well done. Well done. But, yeah, no, I think it's incredibly useful."* (C4)

C4 even inquired about when the system would be available for clinical use:

> *"Yeah, when can we use this? It's Fantastic, it's really good."* (C4)

Output data of the system

Delving deeper and asking about the measures that the system produced, the interviewees were happy with the measures and thought they were all clinically relevant:

> *"Those kind of core skills data presented in the summary are really the key ones that we focus on clinically, and they are some of the ones that people identified as important when we did our research about important outcome domains, that some of those core skills underpin that participation. So that's good."* (C1) *"No, there's nothing there that I think why? Why is that there, no, all of that is everything in there is what we'd want to know."* (C3)

Clinical evaluation of prosthesis users' real-world activity data recorded from the objective sensor system.

Initially when asked if there were any other measures that they thought would improve the clinical relevance, two of the interviewees said that there was nothing they could think of:

> *"There's nothing screaming out there that I'd think, Oh, we're missing."* (C3)

> *"I don't think there's anything else that we would want to know. I don't think so."* (C4)

However, there were other measure that these interviewees mentioned later as the interview progressed, namely force/load and temporal measures.

Regarding force/load measurement, three reasons were given as to why including load through the prosthesis would aid in clinical decision making. The **first** reason provided was that some manufacturers specify that a K4 prosthetic should be given if the user is putting high stress through the prosthesis, e.g.:

> *"K4 you're then saying exhibiting high impact stress or energy levels. So, whether there's any way to capture data related to those factors."* (C1).

The **second** reason provided was to see if the user is putting 50% of their body weight through their prosthesis:

> *"To see actually they are putting 50% of the body weight through the prosthetic."* (C3)

This was viewed as a means to reduce the risk of long-term health issues relating to uneven gait:

> *"In terms of protection of the body and, you know, osteoarthritis and spine. . . "* (C3)

The **third** and final reason given to support a need for load measurement was to see if the user is not only offloading their prosthesis but also has the ability to put additional load through it:

> *"Taking some weight off, you're offloading your leg. Does it allow you to, kind of measure."* (C4)

> *"A lot of them will say that they will carry things. I don't know. Would your sensors show you that, like if you were taking more weight?"* (C4)

Regarding measures relating to wear time and active time were also said to be of interest.

> *"One of the things that I've always found quite interesting is how long someone wears a device."* (C2)

With one reason for this being to better understand how much a user uses their prosthesis, which is a criterium for prescription:

> *"A lot of the criteria for different prescriptions are around full time users so, and yet you can't tell the people that are full time users and then the ones that aren't, and it's very hard to have that discussion if you don't really know."* (C4)

And this interviewee further emphasised that there remains uncertainty about what their patients are saying:

> *"If they say that they wear their leg all day, but actually the reality is they just leave it against a wall all day..."* (C4)

Three of the interviewees stated that active time would be useful not only to judge active periods but also periods of inactivity:

> *"If you could determine standing active time, where they're walking, and how much the leg is used from, from morning to night, then it would help with getting prescriptions right."* (C2)

> *"It's good to know the periods of inactivity, and how long they're wearing the leg for, maybe that'd be quite, that could be quite useful."* (C3)

> *"It's just about that inactivity would be useful, that'd be really useful."* (C4)

In summary, clinicians provided overwhelmingly positive feedback on the system's output, particularly highlighting its clinical relevance and the utility of the data it generates. One clinician even enquired about the system's availability for widespread use. The only additional measures the participants brought up that would be desired clinically were load through the prosthesis, wear time and active time. Active time could already be obtained from the current data, by quantifying stepping periods. The load through the prosthesis could not be obtained with just an IMU, for such an additional sensor would need to be integrated. Measuring the load through the prosthesis was discussed in Chapter 4 as a way of improving walking aid recognition. Load through the prosthesis could then also serve as an accurate way of measuring wear time.

Theme 2: Clinical use

The interviewees further considered how the system could be used clinically, and how it would be helpful for their clinical decision making. Their answers regarding this theme could be categorised into three subthemes: 1) supporting current practice, 2) impact of the system on current practice, and 3) benefit of additional information about patients.

Supporting current practice

Three of the interviewees talked about how the system data could be used to support current practice by proving and justifying clinical decisions, enabling more in-depth conversations on patients' lives, and proving deeper insights into their patients' lives.

> *"I think it actually proves the K levels. It helps you actually have a data driven way of assessing someone's K levels as just now, it's a 20-minute interview. Someone turns up, they tell you something, you look at them, you go, with your gut feeling is that person really doing that or are they not."* (C2)

> *"Yeah, because then it can support clinical decision making."* (C3)

> *"I think it definitely has a place. And I think especially when we are being asked to justify why we're prescribing specific things."* (C4)

One reason why they thought the system data would help is because of the subjectiveness of their current assessments:

> *"It's always good to know extra data with prosthetics, because we're always lacking a little bit behind everywhere else, aren't we, because it's, it's so subjective."* (C3)

C3 expanded on the subjectiveness by adding that this can mean that sometimes the clinician cannot fully believe what their patients are saying:

> *"The patient walks approximately four to five miles a day. We're just going off them saying that. So, we take it with a pinch of salt."* (C3)

C4 gave a reason for why being able to justify prescriptions would be beneficial to a clinic, i.e. because of the cost of prothesis components:

> *"I think it's probably more it would be more frequently used than you can imagine, because you're talking about money here, talking about spending money, saving money, justifying prescriptions."* (C4)

They also commented on how the system could be used practically:

> *"If you had something like this for us to be able to see two weeks prior to them coming in for a review, going to get you to wear this. It would allow us to have a much fuller conversation about why we're going to change their prescription."* (C4)

Impact of the system on current practice
Three of the interviewees further talked about how the system data could change the assessments by facilitating in-depth conversations about their patients' activities:

> *"Often they come in and you say, kind of how do you spend your time? And they go, I like to watch a bit of telly. And then you're like, come on, I need more than that. But it's really hard to get it out. And then, whereas this could help really make those conversations more specific."* (C1)

> *"You can then go into it and say, Well, what, what were you doing on this day that you can, that allowed you to achieve that many and why in the days?"* (C3)

> *"More prompts, more information, to dig a little bit deeper. And it also does affect the ability to justify the prescription."* (C4)

> *"We say to people, you know, do you go up down slopes? That's kind of it. We don't really go into the detail of what type of slope, what kind of surfaces? Is it grass? Is it gravel? Has It got a rail? Whereas this would give you a bit more of a prompt to ask the detail around."* (C4)

C4 gave a bit more context on why current assessments don't always obtain these details whereas having the system data could:

> *"It's not necessarily that we don't want to know these things. I think we just have so many other things to ask them. I think this would allow us to kind of go, okay, well, talk me through, what does that slope look like?"* (C4)

And that the sensor data could make having difficult conversations easier:

> *"I think data, in my mind is always that ability to actually have those conversations without it becoming very personal, you know, feelings and all that kind of stuff. At least, you can kind of use the data to kind of have the kind of more difficult conversation"* (C4)

C1 also highlighted that the questions prompted by the sensor data could also verify the sensor data:

> *"Because when you've got numbers like that, you're, yeah, I would, I would kind of, well, what I'd be saying, like, what happened with all your stair climbing? Like you'd be trying to, you'd be able to pull out some of the inaccuracies in data by verifying it with the patient as best you could."* (C1)

Benefit of additional information about patients

C1 discussed how the system data could give more detail about their patients' lives that could affect clinical decisions:

> *"It would give you an insight into how they live their lives, that's what's often really useful. It can direct those conversations to really unpick what's going on for the patient, and that then affects your clinical decision making, because ultimately, you're trying to match your products to the patient's life."* (C1)

Notably, C4 suggested that the objective system data could potentially be more accurate than insights gained through subjective assessments during a single clinic visit:

> *"I would say it's much, much more accurate than what we currently have by far."* (C4)

And the justification for this was that the system can collect data for two weeks whereas current assessments happen at one specific time:

> *"A lot of the K levels are based on the way that they walk into the building, along with kind of what they say they're doing, the physio team will do an amputee mobility predictor, but that is one point in time. And the problem with that is, if they're having a bad day, then it's bad. If they're having a good day, is good, but it's not consistent."* (C4)

In summary, the clinical experts felt the system data could be used to support current assessments to justify decisions, facilitate conversations that provide more detail about their patients' lives, help with difficult conversations and to obtain more information about their patients' activities. This would make prescriptions more objective and help clinicians be more confident in making clinical decisions.

Theme 3: Presentation of system output

How the system data were presented was also discussed. All of the interviewees gave generally positive feedback on how the data were presented:

> *"I liked the summary, that was really useful."* (C1)

> *"It was nice and simple, was dead straight forward so you can easily print off and stick in a patient's notes. So that was good."* (C2)

> *"It's good to have it broken down"* (C3)

> *"Summary I quite like."* (C4)

C1 and C4 highlighted the percentage of strides above 20% normal cadence as an aspect they both liked:

> *"The other thing I liked was percentage of steps above 20% average cadence, because that was quite useful to think about variable cadence."* (C1) *"You've got the percentage of steps above cadence. I think that is a nice thing to be able to ask a question about, okay, so you walked a wee bit faster. Do you remember what it was you were doing that kind of thing that would allow us to prompt."* (C4)

There were three suggestions made by the participants to improve the way the data is displayed. These were having the data visualised in charts/graphs, having normative data and having percentages instead of just numbers.
Regarding the former, two of the interviewees commented on how the table with the daily data may not get read:

> *"It's not exciting to look at."* (C3)

> *"A lot of numbers in a table like that could be maybe not confusing, but it's just harder to read."* (C4)

C4 also justified this by pointing out that a clinical team might not have the time to decipher the table:

> *"I just know that the clinical team will probably not spend the time that they should do to actually try and understand that."* (C4)

C2 suggested a bell curve for the cadence of different steps:

> *"If you can get a bell curve that would show how much time or how many steps a day on average, they took at different walking speeds. Then it would help."* (C2)

C4 suggested graphs might help but wasn't completely confident with the suggestion:

> *"It's almost like you want to have, a kind of graph, but I don't know if people would be off be put off by that."* (C4)

But they did make it known that it would be beneficial to change how the table is currently displayed:

> *"I feel like it needs to get split up. It's quite a lot to look at."* (C4)

Regarding inclusion of normative data, two of the interviewees thought that typical values for each K level would be helpful:

> *"I suppose normative data is really useful."* (C1)

> *"I don't know if you can do this, is to see, is there a normal value, like, so, like, could you compare?"* (C4)

C4 expanded on this comment by explaining that the clinician might be comparing values against their own measures and not typical prosthetic users:

> *"I think as a clinical team, it would be difficult to know if that was good or bad, because they'll be basing it on their own step counts, which is unrealistic. So, a guy, for example, would know that actually 6500 steps a day is pretty good for an amputee, but I would think a lot of my clinical team might think that's pretty poor."* (C4)

Further regarding the presentation of the data, C4 thought it would be easier to read the daily table if some of the numbers were percentages of daily steps instead of just numbers:

> *"A percentage. Because, quite honestly, if you're doing the table, you can see quite clearly how many steps go up down each day. So maybe it's more of a percentage rather than a proportion, rather than a number."* (C4)

In summary, all the participants liked the summary of the system's data they had been presented with, but some of the participants thought the table of daily activities might not be easy to read, especially for clinicians. Visualizing the data might help, with the use of charts and graphs being suggested, and alternatively percentages of daily steps taken under the various conditions might also make the data easier to understand. Normative data were also suggested, but before provision of such will become possible more data is needed be collected in the real-world.

Theme 4: Comparison of system data with clinical assessments

The interviewees were asked to examine the system's data presented on the lower limb prosthetic users and compare it with the outcomes of the standard clinical assessments. Their thoughts on the comparison was then obtained.

For prosthesis user 1, C1 said that the system data might have changed the prescription for the participant, raising their K level to a K4:

> *"For prosthesis user 1, when I looked at his clinical assessment, I felt like he was a K three, so I thought that matched well. But then, when I looked at his summary data, he seemed really active. So, he was like some days he was right up on 10,000 steps a day, and most days he was hitting that kind of six to 8000 that you might expect for people with a disability. He did a lot of stairs. He did a lot of uneven ground and a lot of slope, and so it made me think about whether he should have actually been a K4 user."* (C1)

They further elaborated:

> *"That kind of data. So, the total number of steps, the steps on different surfaces and during different activities, and the number and the percentage of cadence made me think he should have been a K4 which is interesting, because clinically I wouldn't have put him into K4, so that was that one."* (C1)

Contrary to this, interviewees C2 and C4 both thought the step count for participant 1 was low in relation to the clinical assessment outcomes:

> *"Just seemed to me to be quite low."* (C2)

> *"I think it says 4 to 5 miles a day. And so, I think the sensor data seems to be a wee bit lower than what the patient reports. But I would have thought that four to five miles a day would have been more than 7900 steps."* (C4)

For prosthesis user 2, interviewee C1 thought the system data matched the clinical data:

> *"Participant two, I thought, was pretty straightforward. Like, his clinical assessment and his data, I thought, like were kind of reasonably well matched."* (C1)

Whereas C3 was surprised that the participant was as active as they were from the system data compared to the clinical assessment outcomes:

> *"Participant 2, you wouldn't have necessarily predicted him to be as active because of his condition and his age compared to participant 1, but that, you know, it shows that an 86 year old with a load of comorbidities, hip replacements and everything, can be as active as other people. And I think maybe sometimes we underestimate an aging population. So, I think that's quite useful to see that this man is, obviously a very active man, despite everything else."* (C3)

For prosthesis user 3, the interviewees acknowledged that the participant's fall might have affected the recorded data:

> *"I probably would have put him as a K3, but it's possible that his fall could have made his activity data look worse than it was."* (C1)

Interviewee C3 commented on the same prosthesis user that their step count was low compared to the clinical assessment outcomes:

> *"Participant 3, yeah, reading, what? Reading, you know, type two diabetes, compartment syndrome, after a fall, doesn't smoke, drinks, lives alone, walks 11 kilometres regularly, walks five doesn't run, but cycles a lot, goes to the gym, but then looking at his step count, it's lower than I would have thought it would be."* (C3)

But said that, the breakdown of daily activities was able to help explain this:

> *"Some days his, his total step over 10,000 and then other days, it'd be 1500."* (C3)

And this was perceived to potentially help direct clinical assessments to understand why there are inconsistencies:

> *"It's good to have it broken down, and then you can then go into it and say, well, what, what were you doing on this day that you can, that allowed you to achieve that many and why in the days?"* (C3)

Interviewee C1 also commented on how the system data could change the prescriptions they would give the participants:

> *"I just had the clinical and then actually, I'm looking at the summary data and thinking, oh, my judgments, my assumptions, based on kind of all that clinical data might not actually be that accurate as to how these people live their lives. So, the two together was a really nice combination to help make a decision."* (C1)

They also remarked that it might make them think about what additional support the participants might need in addition to just prosthetic components:

> *"It would also really make me think about the types of prosthesis. Because if he's doing like, he's doing quite a lot of like, if they're doing quite a lot of slope walking, it's like, does he live on a hill? Have we got them on the right knee to be able to cope with the slopes? And what's this slope technique like, it would prompt me to kind of think about some of the skills and make sure that they're good enough for him to be able to do what he needed to do"* (C1)

In summary, the interviewees had different clinical opinions on the participants' data, highlighting the need for an objective system that can support their clinical decision making. The different assessments of the system data gave further arguments for normative data to set benchmarks for each K level to take some bias out of their decision making. The differences highlighted between the clinical assessments and the system data emphasises the subjective nature of the clinical assessment and how patients can have different views on their own activity levels compared to what they actually achieve.

Theme 5: Charging

It was explained to the interviewees that the system used for the data collection needed to be charged by the participants every night. The interviewees were then asked if they thought this could be an issue. None of them thought this should be an issue as most patients are already used to charging equipment at night:

> *"No. So, micro processer prosthetic feet as an example, have to be charged every day."* (C2)

> *"Most people have a mobile phone. They charge that every night. I don't think that's an issue anymore."* (C3)

> *"No, I don't see it being a problem. I think people are very good at charging things up. I don't think it's that much of an issue."* (C4)

Interviewee C3 did give some considerations to the charging that could further eliminate any issues:

> *"Don't think that'd be an issue, as long as it was an easy charging point. We've had issues before with charging points being tiny pins that snap."* (C3)

> *"Now there's a lot of magnetic charging that patients seem to find a lot easier."* (C3)

> *"You just need to make sure that people understand, the cognitive abilities there to understand that. And it's very simple to, you know, it's very clear how to charge it. But I don't, I don't think that's an issue anymore."* (C3)

In summary, none of the interviewees thought a system that needs to be charged at night will be an issue. Consideration should be given to how the system is charged to make it easy and simple to use.

Theme 6: Areas for improvement

There were a couple of negative points that the interviewees made about the data presented, that were in relation to lack of clarity regarding some of the definitions, and also the extent to which activity data are separated in the system's output, both of which are indicating areas of improvement in future work.

Lack of clarity

The definitions of uneven and unstable terrain were highlighted by three of the interviewees as areas of uncertainty within the system's data:

> *"The difference between uneven terrain or unstable terrain, and that, I don't know if I'd know what the difference was between that."* (C1)

> *"I didn't really follow what you, how you differentiate them."* (C2)

> *"What's the difference then, between uneven and unstable?"* (C4)

No definitions were given in the report regarding this, so this is something that should be added in future.

Extent to which activity data are separated.

Two of the interviewees would have preferred if grass, uneven and unstable data were combined:

> *"You could combine all of your grass, uneven terrain and unstable terrain, into uneven terrain, and you could put it all in together."* (C1)

> *"I'm just wondering if there's a need for it to be broken down into that detail, or if you can put them together."* (C4)

C1 gave a justification for this:

> *"There's nothing in the K levels that differentiates the type of uneven surface. It just kind of talks about uneven surfaces."* (C1)

In contradiction to this, interviewee C2 preferred them to be separate:

> *"I prefer them separate."* (C2)

Justifying this with:

> *"The better you break it down, as long as people understand what the differences are, I think you have a better chance of being able to get the best prescription"* (C2)

And they gave an example to further this argument:

> *"If you have someone, for example, who lives in Brighton and is down at the beach all the time, then they're going to have very different requirements than someone who lives in Sheffield and is living in the city centre."* (C2)

In summary, while clinicians were generally positive about the system, they identified several areas where improvements could enhance its usability and clinical relevance. In addition to desiring load measurement and wear and activity times, discussed under output data of the system, definitions need to be added to the system's data, especially for uneven and unstable terrain. Furthermore, there were contrasting thoughts on whether the data needs to be split up into as may categories as it currently has been, and perhaps an option to combine when creating output summaries may be the solution.

6.4 Discussion and Conclusions

The system developed for assessing real-world activity in lower-limb prosthesis users was generally well-received by clinicians, who found its output to be both clinically relevant and useful in supporting their K-level assessments. The measures that have been produced by the system were all deemed to be clinically relevant , with only load through the prosthesis and wear or active time as additional measures suggested that would be desirable.

Notably, load through the prosthesis cannot be measured by the current system. An IMU does not have the capability to measure or accurately estimate load, so an additional sensor would be required in future to produce this outcome. Pressure sensors and load cells have previously been used in lower limb prosthetic user activity studies. Wang et al. [170] and Chen et al. [171] both used just pressure sensing insoles to accurately distinguish between sitting, standing, walking, stair use and stepping over an obstacle. Chen et al. only incorporated one lower limb prosthetic user and 5 able-bodied participants, and only tested the accuracies with 5-fold cross-validation, whilst Wang et al. only tested one participant who was a lower limb prosthetic user. Mai and Commuri [161] used a pressure sensor in the socket to accurately classify stair walking and stair use, but again only for one participant. Liu et al. [172], Xu et al. [173], Liu et al. [125] and Fan and He [162] combined an IMU with load or pressure measurements for locomotion recognition for lower limb prosthetic users. All these studies classified between flat, stair and ramp steps, but with limited participants and all in controlled conditions. Similar studies using IMUs and pressure or load sensors have shown good accuracies at terrain classification for exoskeletons [127][126][107][102][105]. This suggests that additional measurement of load through the prosthesis could improve the terrain classification accuracies that are presently solely IMU-based. Notably, no studies have specifically investigated walking aid recognition from the pressure or load through a prosthesis or insole yet, but Youdas et al. [179] found that participants could offload 25% of body weight using a walking stick and Thies et al. [178] found members of an elderly population would load a walking frame with about 12% body weight. This level of offloading should be detectable with pressure or load data as well, therefore incorporating a load or pressure output may also enhance walking aid use classification.

In contrast to load, active time can already be obtained from the system data and does not require additional instrumentation. The step detection could be used to calculate the length of stepping periods, which would then provide the active time of the user. Prosthesis wear time could also be estimated from the current data by identifying periods of non-activity and assuming the prosthesis is not being worn if there is a long periods where there is no prosthesis. Balkman et al. [42] used a limb presence monitor to measure wear time. Griffiths et al. [169] found that a shank-mounted accelerometer could classify between sitting, standing, stepping and lying. Alternatively, this may also be obtained through load measurements: as a user applies load to a prosthesis when it is being worn, this could serve as a means to quantify prosthesis wear time.

Notably, all the interviewees gave positive feedback on the summary of the participants activities they were provided with, saying that having the data presented in a simple way and easy to read was useful. There were contrasting opinions about the daily activity table, however, some saying that having the daily data would be helpful in seeing how activity changed over the data collection period and could explain some discrepancies between the clinical assessment and the system data, but the interviewees had generally negative views on how it was displayed in that table. One suggestion some of the interviewees had was to visualise the data in some way. A bell curve of cadences was given as one example.

All the interviewees thought the system was clinically relevant and when comparing the system data to the clinical assessment summaries they thought that the sensor data could have changed the K level for some of the participants that they otherwise would not have assigned to them. They all thought the data would be very useful to have to add to their clinical assessment, to justify decisions, prompt in-depth questions on patients' lives, and enable difficult conversations to be had objectively. The use of the data to prompt questions was a positive that the interviewees were very enthusiastic about, as it would allow them to get more information about their patients and, therefore, a greater understanding of their lives, which could help them ensure that their patients are getting the right prescription and support. None of the interviewees thought that a system that needed to be charged every night would be an issue for lower limb prosthetic users. This is as long as the method of charging is easy and straight forward, and magnetic charging ports were suggested as a potential solution for this.

There was debate as to whether to have the activity data split as it currently is, or to combine uneven, grass and unstable terrain all together. The argument for combining these activities into one was that K level descriptions do not distinguish between different uneven terrains and that it would be easier to understand if they are combined. The argument for keeping them split is that it would give the clinicians more detail about their patients' activities. Currently with the data split, the accuracies obtained for leave-some-out were 56.89% and 83.64% for cross-validation, with uneven, grass and unstable combined the accuracies increase to 65.70% for leave-some-out and 87.29% for cross validation. This increase in accuracy is purely down to the high accuracy in the combined uneven class with the accuracies from the other classes unchanged, as can be seen in Figures 6.1 and 6.2. Perhaps providing clinicians with an option to combine, for better accuracies and simpler interpretation, may be the solution.

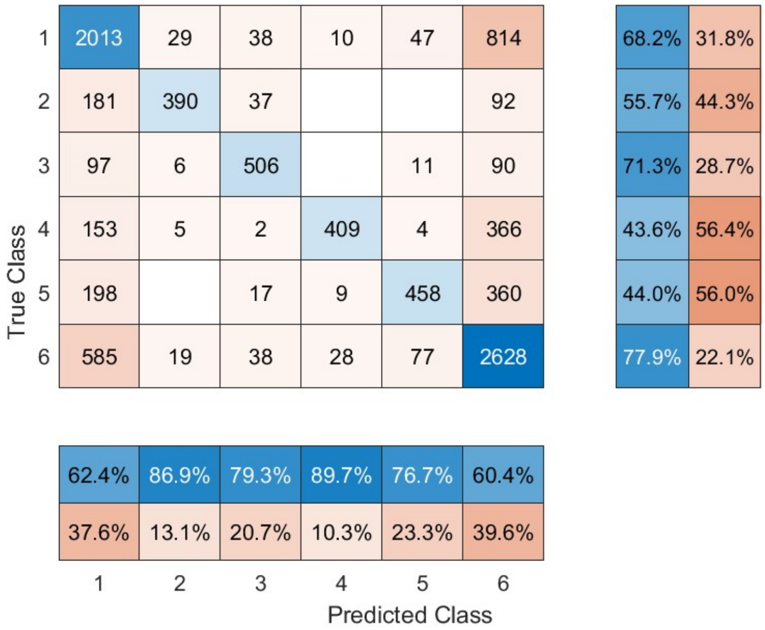

Fig. 6.1 Confusion matrix for leave-some-out: 1-flat, 2-up stairs, 3-down stairs, 4-up slope, 5-down slope, and 6- combined uneven.

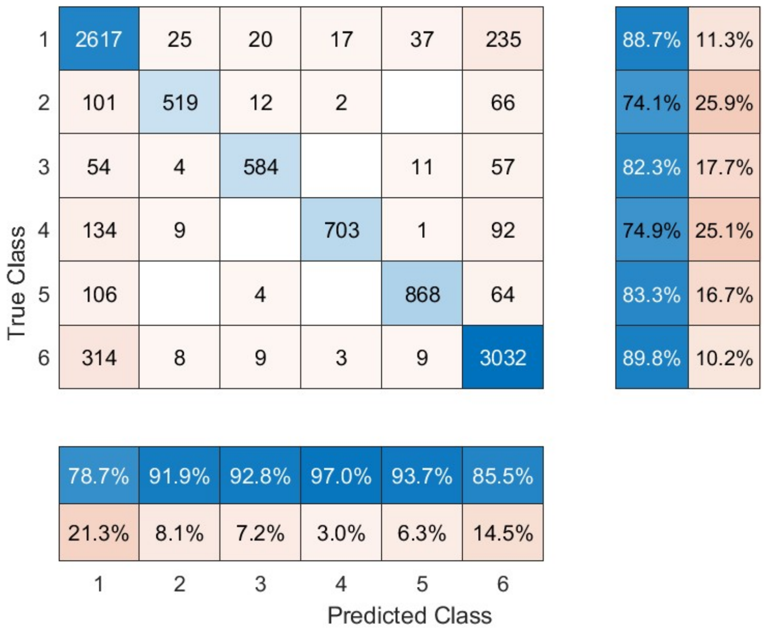

Fig. 6.2 Cross-validation confusion matrix: 1-flat, 2-up stairs, 3-down stairs, 4-up slope, 5-down slope, and 6- combined uneven.

Limitations of the collected data

The accuracies achieved by the models in Chapter 4 were not perfect and although the participants used in this data collection had provided some of the data used to train the models, the terrain they traversed was different to that used to train the models. Hence it cannot be assumed that the same cross-validation accuracies were achieved. Due to this, the terrain classification and walking aid use were only an estimation. However, walking aid use was checked with the participants, and they all said they did not use a walking aid during the collection period, so the low number of steps with a walking aid can be assumed to be representative of the participants. i.e. they may have walked with support of another person or holding on to handrail or furniture on occasion. Participant 3 had a fall on day 3 of the data collection period. This restricted the participant's activities subsequently and could also explain the suggestion of an increase in walking aid use post day 3.

The method for counted steps was assessed and verified in Chapter 4, section 4.2.7, hence the step count and cadence was deemed to be accurate for the collected data.

Even with these limitations, the data were acceptable for the purpose of the study which had the aim to investigate clinicians' thoughts on the system and its output data, presented as examples of what the system is designed to produce. At this time, the data were not being used for any clinical decision making; further training data would be needed from a larger sample to further improve accuracies. Importantly, the interviewees were made aware of these limitations of the data before the interviews.

In conclusion:

The system and the data that are produced were perceived as clinically relevant. Clinicians believed that the system's data on their patients' everyday life would enhance clinical decision making and improve outcomes for their patients, as well as help justify the prescription that they receive. Load through the prosthesis is a measure that clinicians do desire, and which could also improve the classification accuracies, in particular for walking aid use. The summary of the prosthetic users' activities as it had been presented to clinicians was perceived as useful as is, but finding a better way to visualise the daily activities table data could make it easier for clinicians to understand.

Chapter 7

Conclusions and future work

7.1 Background

The aim of the thesis was to develop a clinically-useful, sensor-based system to improve current K level assessments by providing objective information on real-world activity, in particular to address shortfalls in classification at the K2–K3 border. This was to be achieved through a set of 4 objectives:

- To investigate the clinical requirements for the objective system to aid in clinical decision making through interviews that explore clinicians' perceptions regarding shortfalls of current clinical activity assessments for K level assignments, and which objective measures they feel would improve their K level assignment. (Chapter 2)
- To review the literature around sensors and algorithms to inform system design. (Chapter 3)
- To design a system comprised of sensors and algorithms that output these measures, utilizing both, real-world and in-lab data collection. (Chapters 4 and 5)
- To explore clinicians' views regarding the developed system and its outcome measures in the context of real amputee data to assess the clinical benefit the system could provide and identify avenues for further development. (Chapter 6)

K levels are defined by the user's ability to traverse environmental barriers, change cadence and ambulation skill. From the current K level definitions, Table 1.1, a K2 can traverse low level environmental barriers and has limited community ambulation, and a K3 has the ability to traverse most environmental barriers, vary their cadence and conduct activities beyond simple locomotion. It has been shown that if a lower limb prosthetic user is not given an adequate prosthetic that meets their activity needs it could lead to the patient becoming less active and/or not using their prothesis.

This chapter presents a comprehensive discussion of the key findings from this research and will further summarize significance, limitations and directions for future work. It begins with a summary of the results from Chapter 2, which involved interviews with clinical experts to identify current limitations in K-level assessments and opportunities for improvement. This chapter also explores the outcomes of Chapters 3-5, which focused on developing and testing an objective system for terrain and walking aid classification using machine learning models and IMU sensors, trained and tested on data collected in the real-world and the inside the lab. A discussion of its clinical usefulness as explored in Chapter 6 then follows. Finally, limitations and directions for future work will be discussed.

7.2 Thesis findings

7.2.1 Investigating the clinical requirements for the objective system through interviews (Chapter 2)

Interviews were carried out with clinical experts to gain an understanding of current clinical shortfalls in relation to lower limb prosthetic prescriptions and which objective measure if recorded in the real-world would have clinical benefit. The interviews found that the assessments of K levels are mainly based on a conversation between the clinician and their patient, with no set guidelines on how these conversations should take place. The clinician then has to make a subjective decision on the patient's K level which determines the prosthesis components the patient will be prescribed. The clinicians identified that this, in conjunction with the vagueness of the K level descriptions, is the main source of problems with current prosthetic prescriptions. These results are reinforced by findings from Jamieson [40] that found that patients do not keep their clinicians updated on their activity levels except when needed and Limb et al. [41] and Balkman et al. [42] that showed that self-reported measures of activity are not accurate when compared to objective measures. The clinical experts highlighted step count, cadence, walking aid use and the terrain a prosthetic user traverses as the key objective measures that would help with clinical decision making. Terrain was also found by Jamieson [40] to be a measure clinicians desired to know for prosthetic component prescriptions. A few studies had previously investigated step count and cadence against K level and found some alignment but had over 11% misalignment between these measures and the assigned K level. This highlights the need for terrain and walking aid use as measures to be combined with step count and cadence to improve prosthetic component prescriptions. It was also stated that longer term monitoring would be beneficial, because activity levels can differ on a daily basis, as found by Knols et al. [43] where timed walked tests did not correlate to long term step count data. This was also shown in Chapter 6 where daily step count varied for all participants, the lowest variation to mean daily steps was 25% and the highest was 160%. For acceptance from prosthetic users, the system will have to be mounted to the prosthesis only. The importance of getting the assessments correct was highlighted by Agrawal et al. [37] that found that the components a prosthetic user uses can have a large effect on their activity levels.

7.2.2 Reviewing the literature around sensors and algorithms to inform system design (Chapter 3)

A review of the current literature provided a number of useful insights for system design: it showed that terrain classification has been previously investigated but very few studies incorporated terrain outside of a laboratory setting, recruited a substantial number of lower limb prosthetic user participants, investigated complex terrain, for example uneven terrain or unstable terrain, or used sensors only attached to a prosthesis. Machine learning models have been used and shown to be able to classify terrain from wearable sensor data but there is no consensus on the best techniques or methods for its implementation. Very few studies have explored walking aid recognition using wearable sensors and none have studied walking aid recognition for lower limb prosthetic users. The studies that have investigated walking aid recognition using wearable sensors have used wrist worn sensors which would reduce the acceptability of the system for the prosthetic users. The insights gained from the literature review led to a list of specifications that were then taken forward to system design, namely that IMUs were to be used to capture activity data at 100 Hz, which were then to be low pass filtered and normalized, and furthermore that KNN, LR, SVM, RF and LSTM algorithms would be compared for terrain classification whilst the same plus LR algorithms would need to be compared for walking aid recognition, and, finally, use of a variety of time-based and stride-based windows as well as analysing individual datapoints would need to be investigated.

7.2.3 System design (Chapters 4 and 5)

Chapters 4 and 5 then aimed to develop an objective system to classify terrain and walking aid use that would be robust so it could be used to aid clinical decision making in the real-world. In doing so, gaps in the research previously found around algorithms and sensors were filled.

In **Chapter 4**, data were collected of lower limb prosthetic users walking outdoors over different terrain (flat ground, stairs, ramps, grass, gravel and uneven terrain) with and without a walking aid using four IMUs attached to different parts of the participant's body and prosthesis. Twenty participants were recruited, eleven unilateral transtibial, eight unilateral transfemoral and one bilateral with one transtibial and transfemoral prosthetic. The data were collected at either the University of Salford or at a location convenient for the participant, and this variety should have helped making the model be more robust.

For terrain classification, four machine learning algorithms were compared (KNN, RF, SVM and LSTM). Other aspects of the model were also investigated to see how the IMU location, sampling rate, windowing method, variable inclusion and normalisation technique affected the accuracies of the classifiers. It was also investigated whether splitting the data for transtibial and transfemoral prosthetics and whether or not a walking aid was being used could help improve the accuracies. It was found that a single prosthetic-shank mounted IMU was only needed to produce high terrain classification accuracies. Splitting the data into windows that contain a whole stride was the best performing windowing method, which is a consideration that future similar studies should bear in mind as this method is not the most utilised in the literature. It was found that windows of just 10 datapoints could be employed to maintain a high accuracy, which means that a sampling rate of 20Hz is needed. Accelerations, gyroscope, magnetometer and free accelerations were needed to maintain high accuracies, whereas velocities did not affect accuracies. Normalising the data per person and keeping the raw data upheld the high accuracies, but normalizing per stride did not. Splitting the data by prosthetic type or whether a walking aid was used did not affect the accuracies. A RF model produced the highest leave-some-out accuracy at 56.89% and a KNN model produced the highest cross-validation accuracy at 85.71%. the The KNN model produced a low leave-some-out accuracy of 45.67% whereas the RF model produced an 83.64% cross-validation accuracy, ; due to this the RF model was taken forward to the subsequent chapters. The leave-some-out accuracy might seem low but the most comparable study by Jamieson [40] produced a leave-some-out accuracy of 56.68% when only categorising the terrain as either flat, stair ascent, stair descent, ramp ascent and ramp descent, ; with the same terrain grouping the RF model produces a leave-some-out accuracy of 79.02%. The cross-validation accuracy was also higher than Jamieson, with the RF model producing an accuracy of 90.29% and Jamieson achieving 78.46%;, the KNN model out performed the RF model again and produced an accuracy of 93.32%. These accuracies are comparable to other studies that classified limited terrain in laboratory conditions [121][120][107], and the cross-validation accuracies showed the accuracies that could be achieved. Collecting data on a variety of terrains will help to make the models more robust but the limited number of strides on each variation in the current data could be lowering the leave-some-out accuracies; this highlights the need for a larger dataset to be collected incorporating different variation in the terrain. The variation in lower limb prosthetic [197] could also be contributing to the lower leave-some-out accuracies. Although the dataset contained data from more lower limb prosthetic users than had been collected in previous studies (9 by Du et al. [130]), a larger dataset would be needed to achieve high accuracies for clinical use.

For the walking aid recognition five machine learning algorithms were compared (KNN, RF, SVM, LSTM and LR). LR was incorporated in the walking aid recognition development but not the terrain classification as LR works best with binary classification. The same aspects of the models as for the terrain classification were also investigated (IMU location, sampling rate, windowing method, variable inclusion and normalisation technique). It was also investigated if analysing the data separately for different terrains and prosthetic types would affect the accuracies. It was found that the IMU on the trunk produced the best accuracies at 67.95% leave-some-out and 84.84% cross-validation, but the prosthetic shank IMU produced accuracies that were only slightly lower with the same conditions of 64.39% leave-some-out and 82.21% cross-validation. Due to the requirement that the system only mounted to the prosthesis it was deemed that just the prosthetic shank IMU would be used as the accuracies were similar to the trunk IMU. Splitting the data into windows containing a whole stride again produced the best accuracies and it was found that reducing the window to 10 datapoints did not affect the accuracies. Free acceleration and magnetometer data did not improve the accuracies, whereas accelerations, gyroscope and cadence data were critical. As with the terrain classification, normalising per person and raw data combined produced the best accuracies and splitting the data into prosthetic type or terrain did not affect the accuracies. A SVM model produced the highest accuracies with 65.03% leave-some-out and 84.42% cross-validation. Considering that 61% of the strides in the data were taken without a walking aid, the leave-some-out accuracy is low. The confusion matrix, Figure 7.1, shows that less than 50% of the strides without a walking aid were correctly classified, which suggests that the model overclassified to without a walking aid, but even with this, strides without a walking aid were misclassified 26.1% of the time. The cross-validation accuracy theoretically shows what could be achieved if all variation is accounted for, but as seen in the confusion matrix, Figure 7.2, strides with a walking aid were misclassified 19.8% of the time and without 12.9% of the time. which might not be acceptable for clinical use. As walking aids are used to increase balance and offload, measuring the load through the prosthesis could be a more accurate method of classifying walking aid use.

Fig. 7.1 Confusion matrix for walking aid recognition Leave-some-out SVM optimised trial. 1-without a walking aid 2-with a walking aid.

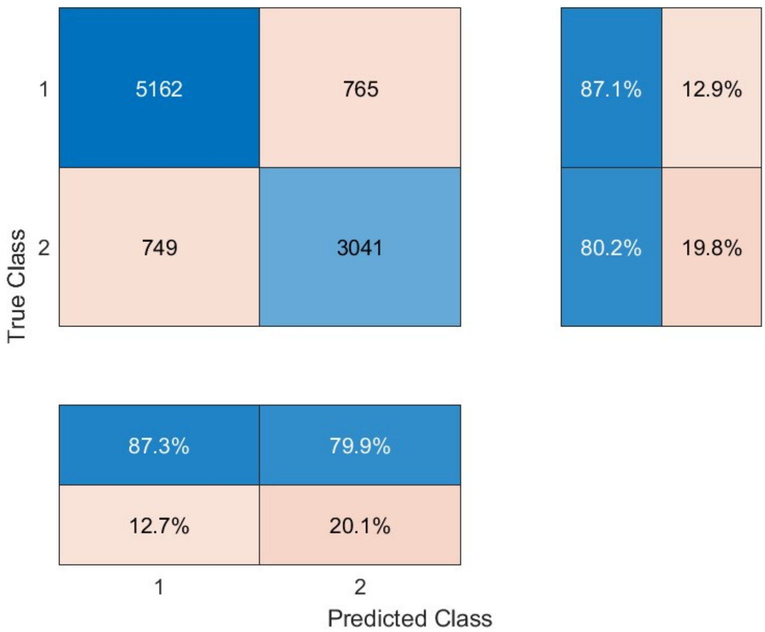

Fig. 7.2 Confusion matrix for walking aid recognition Cross-validation SVM optimised trial. 1-without a walking aid 2-with a walking aid

The location of the IMU on the prosthetic shank was not fixed for the data collection. The IMU was attached to the pylon and where possible aimed to be attached in the middle of the shank, but on some of the participants prothesis this was not possible, so the location did vary. To investigate if this could be affecting the accuracies and if there is an ideal location, **Chapter 5** looked at the variance of features between classes (terrains and walking aid use), the theory being that a larger variance will increase the classification accuracies. The variance was examined in a couple of ways, first looking at the classification models and identifying the most important features then seeing how these features changed over the shank, and comparing classes individually to see which features produce the largest variance and which shank location produce large variances. Stereophotogrammetry data were used to create virtual sensor signals and the virtual signals were looking at how acceleration changed along the shank, as gyroscope and orientation data should be consistent. Three locations were chosen, towards the knee, mid shank and towards the ankle. The stereophotogrammetry data were captured in a Qualisys motion capture laboratory which restricted the terrain that could be tested. Flat ground, stair use, ramp walking and cobble stones were the terrains used and data were collected on each with and without a walking aid. Ten lower limb prosthetic users were recruited for this part of the study, six transtibial and four transfemoral. For terrain classification, all aspects of the accelerations were deemed important, although acceleration normalised had more influence than raw accelerations. The ankle accelerations produced higher variance for more terrain comparisons, but when data for the three locations were put into the classification model there was no significant difference between the locations. It was found that a consistent location did improve the accuracies, however. For walking aid recognition the x component (vertical) of the acceleration had a smaller influence over the classifier. There was no difference in the highest variance between the three locations, but the knee signals did produce higher classification accuracies, and again using data from different locations reduced the accuracies.

7.2.4 Clinical evaluation of the system (Chapter 6)

A final study was undertaken to evaluate the system that had been developed. This was achieved by obtaining clinical experts' opinions and thoughts on the system. Data were collected from three lower limb prosthetic users (two transtibial and one transfemoral) over two weeks, and the data were classified with the models created in the preceding work. The prosthetic users also had a clinical assessment to be assigned a K level as per standard clinical practice. Interviews were then conducted with four clinical experts. This was the first study to evaluate a real-world, long-term data collection system for lower limb prosthetic users, designed specifically to assist clinicians in K-level assessments through objective sensor-based data. The clinical experts were presented with a report on each participant, that summarised the clinical assessment and the assigned K level, and further presented the system data in a summary and a daily breakdown, before the interviews took place. All the interviewees gave positive feedback on the system and viewed the data as clinically relevant. The clinicians all thought that the data would be beneficial for supporting K level assessments and could improve clinical outcomes for patients. They all also liked how the data summary was presented. The only additional data that they thought could also be clinically beneficial was load through the prosthesis and wear/active time. The only other improvement that was suggested was to have the daily breakdown data easier to read: the data had been presented to the clinical experts in a table, which some of them felt could be too hard to read for clinicians and visualising the data might make it easier. Another suggestion was to reduce the amount of terrain categories by combining uneven, grass and unstable to one group. If this was to be implemented, the current classification accuracies would improve to 65.70% leave-some-out and 87.29% cross-validation.

7.3 Limitations

There were a several limitations to this research that will be discussed in this section. These limitations include participant recruitment, time limitations, and data analysis.

7.3.1 Recruitment of participants

Although more lower limb prosthetic users were recruited for the data collection in Chapter 4 than in any previous similar research, participants were only recruited from a few sources. Eleven of the twenty participants were professional patients at the University of Salford, one was a professional patient at the University of Strathclyde, two through ManFit (a prosthetic users fitness initiative in Greater Manchester) and six through Portsmouth Enablement Centre NHS. This meant that the data collection only took place in Greater Manchester and South Hampshire, which could limit the variation in terrain that the models have been trained on. Before the system could be ready to be used clinically, data would have to be collected from more varied locations and terrains to ensure the robustness of the model.

7.3.2 Time limitations

Several factors contributed to delays in aspects of the research that resulted in time limitations on the later sections of the research. Recruitment for the data collection in Chapter 4 was a large contributor to these delays. University of Salford and NHS ethical approval had been obtained in July 2022, but the first participant was not able to be recruited until November of 2022. Initially, recruitment was intended to take place through the Manchester University NHS Foundation Trust Specialised Ability Centre, but after three months the centre was unable to recruit any participants so alternative avenues of recruitment had to be pursued. To reduce the burden on participants, a change to the study procedure was made meaning that data could be collected at a location convenient for the participant. This did have the benefit that data were collected from more varied terrains. These delays meant, however, that the data collection continued until March 2024 which reduced the time for model development and subsequent clinical evaluation.

7.3.3 Data analysis

There were a few areas of the data analysis that had to be reduced due to time restrictions. One of these was testing different aspects of the model. In Chapter 3, the IMU position, windowing method, sampling rate, variables and normalisation were individually tested but due to time restrictions the interactions between them were not explored. The processing time for some of the trial iterations were over 24 hours, which when there were five iterations per model for each trial made running all combination of model aspects not feasible. This means that from this research it can not be said that reducing the sampling rate to 20Hz is acceptable for all windowing methods, but only for windows of a stride that contains 10 datapoints for prosthetic shank IMU data. The time limitations also meant that not all techniques could be tested, one example of this is sliding windows. Sliding windows is a time-based windowing technique where the windows overlap and has been used in previous studies but did not show to produce higher accuracies than stride-based or time-based non-sliding windows, although comparisons between the methods on similar data has not been conducted. Future development could look at sliding windows to compare them against stride-based windows.

The data collected for Chapter 4 were collected in real-world environments but under test conditions. The participants walked on each terrain in separated trials. Gait can change between test conditions and free real-world walking [198]. It would be ideal if data could be collected in a free real-world. For this, a system to record the terrain on every step would have to be investigated. This could be new research, following the participant and video recording their movements for a period of time, with the system recording data and then labelling the data using the video recordings. This may provide more data and more natural gait data but will increase processing time to label the data. A visual system attached to the participant or their prosthesis that can identify terrain, for examples such as the camera system from Qian et al. [129] or the laser system Luo et al. [115] investigated, could be used to label data and would reduce processing time, and all the participants would freely walk to collect data.

There are questions about how to classify different terrains. One of these questions is when does a terrain fit into a certain category. One example of this is slopes, for the data collection a slope was considered if it was steeper than 1:20, which was informed by the UK building regulations [176]. This does not mean that only slopes steeper than 1:20 will be classified as a slope by the model, the steepness a slope will need to be to be classified is not known. It is also not known if a much steeper slope will be classified, but McIntosh et al. [199] and Strutzenburger et al. [200] did not find a significant different between gait parameters for slopes between 5° and 12°, but neither of these study incorporated prosthetic users. Uneven terrain is another area for debate. For the data collection there were no strict criteria for uneven terrain; it was decided on by the researcher conducting the trial. This could result in an overlap between terrain considered flat and terrain considered uneven. Thomas et al. [201] investigated objective measures to assess the complexity of different terrains, but these measures required multiple photographs to be taken from different angles to be taken for each terrain which would not be feasible in a free world condition. It would also increase the data processing time.

Active time was not included in the reports in Chapter 6. Active time could have been calculated from the recorded data if each stepping bout is taken as an active period. Wear time was mentioned by some clinical experts in Chapter 2 as a measure that could help clinical decision making but not active time. Active time however was highlighted by the clinical experts in Chapter 6 as an additional measure they think would have a clinical benefit.

Load through the prosthesis was another measure that the clinical experts in Chapter 6 mentioned would be of clinical benefit. It is a measure that has been utilised in previous studies to aid in terrain classification [125][97][107] and wear time [71]. It is also a measure that could aid in walking aid recognition by detecting offloading, although this has not been tested previously. Load through the prosthesis was not included as part of the system in this research. IMUs alone had previously been used to classify terrain, and load through the prosthesis was not mentioned by the clinical experts in Chapter 2, due to this only IMUs were used in Chapters 4-6. Another factor would be that without altering the participants prosthesis the only method of recording load through a prosthesis that has been previously studied is measuring the pressure the prosthetic foot exerts on a shoe with a pressure sensing insole. For long-term real-world use this would mean no data will be recorded when the participant is not wearing shoes and also will require the participant to move the pressure sensing insole to the shoe they are wearing if they change shoes. The insole would also need to be synchronised with the IMU in the morning as the IMU does not record data while it is being charged. This would increase the burden on the participant which may not be acceptable for clinical use.

7.4 Success of this Thesis

Meeting Objective 1: "To investigate the clinical requirements for the objective system to aid in clinical decision." - The work in Chapter 2 was undertaken to understand the current clinical shortfalls, which were mainly due to the subjectivity of the assessments and the interpretation of the vague K level definitions. The objective measures that could have clinical benefit were also obtained, and they were the cadence and cadence variations a prosthetic user walks at, the terrain they are able to traverse, whether they use a walking aid and their step count.

Meeting Objective 2: "Review the literature around sensors and algorithms to inform system design." - Chapter 3 reviewed the literature concerned with sensors and algorithms, and produces insights and specification for the subsequent system design.

Meeting Objective 3: "Design a system comprised of sensors and algorithms that outputs the clinically-relevant measures" - Chapters 4 developed an IMU system that is able to classify terrain, count steps and measure cadence using a single prosthetic shank mounted IMU. Walking aid recognition was not proven to be able to be classified from a single prosthesis mounted IMU to the accuracies that would be needed for clinical use, however, load measures were identified as a solution to be investigated in future work.

Building on the successes of Chapter 4, Chapter 5 refined Objective 3 by optimising IMU placement. It was discovered that consistent IMU location is required to produce higher classification accuracies and that measuring acceleration towards the ankle of the prosthetic shank will improve terrain classification accuracies.
Meeting Objective 4: "To explore clinicians' views regarding the developed system and its outcome measures" - Chapter 6 had clinical experts evaluate the system developed in the preceding work, to assess the clinical benefits it could have and avenues for further development. This was done by using the new system to collect real-world long-term data on lower limb prosthetic users and presenting the processed data to clinical experts. The

data recorded were deemed to be of clinical benefit. The further development identified was to incorporate load through the prosthesis and wear time to the system.

7.5 Strength and Significance

One strength of this PhD lay in the qualitative research that aimed to obtain clinical experts views to not only direct the research to ensure it is clinically relevant but also to evaluate the system post-development. To the authors knowledge no previous similar research has taken this approach. The post-development evaluation highlights the benefits of conducting clinically directed research, with the feedback on the system that the clinical experts provided being positive, finding the system clinically relevant, it was also highlighted that the system data could affect K level decisions and therefore improve patient outcomes. Their involvement also obtained further development ideas that would not have been considered if it had not been for the post-development evaluation.

Another strength lay in the sample size. Previously the largest sample of lower limb prosthetic users that had been studies in similar research was nine [130], whilst in this research twenty were able to be recruited. Although the recruitment pathways were limited, this emphasises the success of the recruitment strategy. Collecting data in a location that is convenient for the participant made it easier for participants to take part in the study, which increased recruitment.

Furthermore, the terrain classification accuracies achieved from a single prosthesis mounted IMU with data collected on real-world terrain are comparable to studies using much more complex systems in laboratory conditions, this will reduce cost and burden on prosthetic users. The use of real-world terrain will make the models more robust and the accuracies more reliable. This pushed the system closer to clinical readiness than previous research. It was also identified that keeping the location of the IMU would increase locomotion activity classification accuracies, which has not previously been investigated. The research also showed that a single prosthesis mounted IMU can not classify walking aid use to an acceptable accuracy for clinical use, but identified load measures as a solution to be pursued.

Finally, a few aspects of the machine learning models were compared in ways that had not been done previously on similar data. The windowing method is one of these; previously studies had compared window size but not different methods. This research showed that splitting data into strides produced the highest accuracies. Only thirteen out of ninety-one reviewed in Chapter 3 used a stride-base window. It was also found that the sampling rate could be reduced to 20Hz without affecting the high accuracies. Another finding from this research was that a consistent location for the IMU will result in higher classification accuracies. Moreover, whilst there was not a particular location that significantly produced higher classification accuracies, there was higher variance between terrain accelerations towards the ankle.

7.6 Recommendation for future Work

To build on the success of thesis future work is planned to improve the system and progress towards the system being used clinically to improve outcomes for lower limb prosthetic users. The planned future work includes incorporating load through the prosthesis into the system, further data collection to improve model accuracies and robustness, and conducting a reliability study to see how the system affects K level assignments.

Load through the prosthesis is a measure that has been highlighted in a number of areas of this thesis. It will give clinics data on the amount of stress their patients are putting through their prosthesis which will give justification for certain prosthesis components. It could also potentially improve terrain and walking aid classification accuracies. Another use for this data would be to calculate wear time, which is the other measure clinical experts thought would be of benefit. Initially it is planned to obtain these data through pressure sensing insoles to gather an understanding on how these data could be used and their benefit to the system. If it is deemed a benefit to the system, then further development will investigate how these data could be gathered in way that is acceptable to lower limb prosthetic users.

Although this research was able to collect data from twenty lower limb prosthetic users on a variety of real-world terrains, the leave-some-out accuracies indicated that more data is needed before the system could be used clinically. The classification models have to be robust enough to be able to classify the variety of terrain that a lower limb prosthetic user could traverse. They will also have to be able to incorporate the variance in gait for different lower limb prosthetic users. These data will also ideally be collected in free-world conditions and not test-conditions to capture a more natural gait.

Finally, to understand how the system data affects the consistency of K level assignments from different clinicians, a reliability study is needed. The study would involve two groups of clinicians from multiple clinics, every clinician in each group will assess a pool of lower limb prosthetic users to assign each a K level. Specifically, one of the groups of clinicians will only use their current clinical assessments to decide the prosthetic users K levels, whilst the other clinician group will be given data collected by the system on each prosthetic user over a two-week period to support their K level assessments. The intra-reliability between the assigned K levels will be assessed. If there is less variance in the assigned K levels for the group that were aided by the system data, it will show how the system can help to standardise K level assignment between clinicians and clinics. Qualitative data would also need to be obtained from the prosthetic users and clinicians to help lead further development and to further explore the clinical benefit of the system, plus the acceptance from patients.

References

[1] Ottobock, "Kintrol | safe mobility.," 2022. https://www.ottobock.com/en-gb/product/VS4~50_B.

[2] Ottobock, "Terion | confidence with every step," 2022. https://www.ottobock.com/en-gb/product/1C10.

[3] Ottobock, "Taleo harmony | ready for everyday life.," 2022. https://www.ottobock.com/en-gb/product/1C52.

[4] Össur, "Össur prosthetic knees and legs | prosthetic leg | Össur uk," 2022. https://www.ossur.com/en-gb/prosthetics/knees.

[5] E. M. Balk, A. Gazula, G. Markozannes, H. J. Kimmel, I. J. Saldanha, L. J. Resnik, and T. A. Trikalinos, *AHRQ Comparative Effectiveness Reviews*. Agency for Healthcare Research and Quality (US), 2018.

[6] ISPO-UK, "Amputee reported outcome measures," 2021. https://amprom.uk.

[7] E. Condie, H. Scott, and S. Treweek, "Lower limb prosthetic outcome measures: A review of the literature 1995 to 2005," *JPO: Journal of Prosthetics and Orthotics*, vol. 18, no. 6, 2006.

[8] Limbs4life, "Lower limb feet and knees," 2022. https://www.limbs4life.org.au/prosthetics/directory/lower-limb.

[9] W. H. Organization *et al.*, "Who-standards-for-prosthetic-and-orthotics-services-provision part 1," *Geneva, Switzerland*, p. 53, 2017.

[10] K. Ziegler-Graham, E. J. MacKenzie, P. L. Ephraim, T. G. Travison, and R. Brookmeyer, "Estimating the prevalence of limb loss in the united states: 2005 to 2050," *Arch Phys Med Rehabil*, vol. 89, no. 3, pp. 422–9, 2008.

[11] M. Caruso and S. Harrington, "Prevalence of limb loss and limb difference in the united states: implications for public policy," *White Paper. Avalere and Amputation Coalition*, 2024.

[12] M. Twiste, "Limbless statistics annual report: 2011/2012," 2015.

[13] International-Diabetes-Federation, "International diabetes federation. idf diabetes atlas, 10th edn. brussels, belgium: International diabetes federation, 2021," 2021.

[14] E. P. Weledji and P. Fokam, "Treatment of the diabetic foot - to amputate or not?," *BMC Surg*, vol. 14, p. 83, 2014.

[15] International-Diabetes-Federation, "Idf clinical practice recommendations for managing type 2 diabetes in primary care," 2019.

[16] J. B. Bussmann, E. A. Grootscholten, and H. J. Stam, "Daily physical activity and heart rate response in people with a unilateral transtibial amputation for vascular disease," *Archives of Physical Medicine and Rehabilitation*, vol. 85, no. 2, pp. 240–244, 2004.

[17] J. Langford, M. P. Dillon, C. L. Granger, and C. Barr, "Physical activity participation amongst individuals with lower limb amputation," *Disability and Rehabilitation*, vol. 41, no. 9, pp. 1063–1070, 2019. doi: 10.1080/09638288.2017.1422031.

[18] W.-H.-O. and others, "The world health report 2002: reducing risks, promoting healthy life," *World Health Organization*, 2002.

[19] B. R. Smidt KP, "Prosthetics in orthopedics," *StatPearls*, 2023.

[20] B. Dupes, "Prosthetic knee systems. amputee coalition," *amputee coalition*, 2014. https://www.amputee-coalition.org/resources/prosthetic-knee-systems/.

[21] The-War-Amps, "Prosthetic knees for adults," 2024. https://www.waramps.ca/pdf/english-site/ways-we-help/artificial-limbs-and-devices/lower-limb/knees-for-adults.pdf.

[22] A. Holz, "Prosthetic leg fit: Types of knee prosthesis for leg amputations," 2022. https://propelphysiotherapy.com/amputation/prosthetic-legs-types-of-knee-prosthesis/#:~:text=Types%20of%20Knee%20Prostheses,-Prosthetic%20knees%20allow&text=Prosthetic%20knees%20are%20divided%20into,%2Daxis%2C%20and%20polycentric%20knees.

[23] Össur, "Power knee user guide," 2024. chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://media.ossur.com/ossur-dam/image/upload/product-documents-global/Power-Knee_en_za.pdf.

[24] Össur, "Locking knee catalog," 2024. chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://media.ossur.com/ossur-dam/image/upload/product-documents-global/Locking_Knee_gb_en_PN20045.pdf.

[25] Cleveland-Clinic, "Prosthetic foot," 07/08/2024 2024. https://my.clevelandclinic.org/health/treatments/prosthetic-foot.

[26] Amputee-Coalition, "Prosthetic feet," 08/2016 2016. https://www.amputee-coalition.org/resources/prosthetic-feet/.

[27] Physiopedia-contributors, "Prosthetic feet," 18 August 2022 2022. https://www.physio-pedia.com/index.php?title=Prosthetic_Feet&oldid=314534.

[28] J. Gardiner, A. Z. Bari, L. Kenney, M. Twiste, D. Moser, S. Zahedi, and D. Howard, "Performance of optimized prosthetic ankle designs that are based on a hydraulic variable displacement actuator (vda)," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 12, pp. 2418–2426, 2017.

[29] E. C. Baars, E. Schrier, P. U. Dijkstra, and J. H. B. Geertzen, "Prosthesis satisfaction in lower limb amputees: A systematic review of associated factors and questionnaires," *Medicine (Baltimore)*, vol. 97, no. 39, p. e12296, 2018.

[30] Amputee-Coalition, "Do you know your k-level? an interview with interim president ceo sue stout about the importance of amputees knowing their k-levels," *inMotion*, vol. 23, no. 5, 2013.

[31] H.-C.-F.-A. (HCFA), "Hcfa common procedure coding system (hcpcs)," 2001.

[32] B. J. Hafner, "Clinical prescription and use of prosthetic foot and ankle mechanisms: A review of the literature," *JPO: Journal of Prosthetics and Orthotics*, vol. 17, no. 4, pp. S5–S11, 2005.

[33] H.-P.-H.-C. Tufts-Health-Plan, "Medical necessity guidelines: Lower limb prostheses," report, Tufts Health Plan Harvard Pilgrim Health Care, 1/06/2022 2022. https://tuftshealthplan.com/documents/providers/guidelines/medical-necessity-guidelines/lower-limb-pros.

[34] ShelteringArms®, "Lower limb prosthesis 101: Knowledge is power," 2021. https://www.shelteringarms.com/rehablog/lower-limb-prosthesis-101/.

[35] Pennmedicine.org, "University of pennsylvania health system | penn medicine," 2018. https://www.pennmedicine.org/.

[36] V. S. Nelson, K. M. Flood, P. R. Bryant, M. E. Huang, P. F. Pasquina, and T. L. Roberts, "Limb deficiency and prosthetic management. 1. decision making in prosthetic prescription and management," *Arch Phys Med Rehabil*, vol. 87, no. 3 Suppl 1, pp. S3–9, 2006.

[37] V. Agrawal, R. S. Gailey, I. A. Gaunaurd, C. O'Toole, A. Finnieston, and R. Tolchin, "Comparison of four different categories of prosthetic feet during ramp ambulation in unilateral transtibial amputees," *Prosthetics and Orthotics International*, vol. 39, no. 5, pp. 380–389, 2015.

[38] "K level selection reliability? variation in long term activity monitoring (over 8 months)."

[39] A. G. Jamieson, L. Murray, and A. Buis, "The use of physical activity outcomes in rehabilitation interventions for lower limb amputees: a systematic review," *Can Prosthet Orthot J*, vol. 3, no. 1, p. 33931, 2020.

[40] A. Jamieson, *Development of a clinically-targeted human activity recognition system to aid the prosthetic rehabilitation of individuals with lower limb amputation in free living conditions*. Thesis, University of Strathclyde, 2021.

[41] E. Limb, S. Ahmad, D. Cook, S. Kerry, U. Ekelund, P. Whincup, C. Furness, and T. Harris, "P134 does physical activity based on self-completed questionnaire assessment give the same answers as objective accelerometry in physical activity trials? results from the pace-up trial," *Journal of Epidemiology and Community Health*, vol. 70, no. Suppl 1, pp. A112–A113, 2016.

[42] G. S. Balkman, A. C. Vamos, J. E. Sanders, B. G. Larsen, and B. J. Hafner, "Prosthetists' perceptions of information obtained from a lower-limb prosthesis monitoring system: A pilot study," *JPO: Journal of Prosthetics and Orthotics*, vol. 31, no. 2, pp. 112–120, 2019.

[43] R. H. Knols, E. D. De Bruin, D. Uebelhart, and N. K. Aaronson, "The relationship between ambulatory step activity, self-reported physical functioning and standardised timed walking in patients with haematological malignancies," *Disability and Rehabilitation*, vol. 32, no. 22, pp. 1819–1826, 2010.

[44] ISPO, "Lead and compass: Defining outcome measures and a core dataset for the lower limb prosthetics sector," report, International Society of Prosthetics and Orthotics, September 2021 2021. https://www.ispoint.org/lead-compass/.

[45] R. Hanspal, K. Fisher, and R. Nieveen, "Prosthetic socket fit comfort score," *Disability and Rehabilitation*, vol. 25, no. 22, pp. 1278–1280, 2003.

[46] Physiopedia, "Trinity amputation and prosthesis experiences scales," 2022. https://www.physio-pedia.com/Trinity_Amputation_and_Prosthesis_Experiences_Scales#:~:text=Developed%20by%20the%20Department%20of.

[47] D. A. Boone and K. L. Coleman, "Use of the prosthesis evaluation questionnaire (peq)," *JPO: Journal of Prosthetics and Orthotics*, vol. 18, no. 6, 2006.

[48] P. Gallagher, O. Horgan, F. Franchignoni, A. Giordano, and M. MacLachlan, "Body image in people with lower-limb amputation: A rasch analysis of the amputee body image scale," *American Journal of Physical Medicine Rehabilitation*, vol. 86, no. 3, pp. 205–215, 2007.

[49] AustPAR, "Austpar - measurement tools," 2022. http://www.austpar.com/portals/gait/tools.php.

[50] Shirley-Ryan-AbilityLab, "Frenchay activities index," 2019. https://www.sralab.org/rehabilitation-measures/frenchay-activities-index.

[51] C. Gauthier-Gagnon and M. C. Grisé, "Prosthetic profile of the amputee questionnaire: validity and reliability," *Arch Phys Med Rehabil*, vol. 75, no. 12, pp. 1309–14, 1994.

[52] M. Devlin, T. Pauley, K. Head, and S. Garfinkel, "Houghton scale of prosthetic use in people with lower-extremity amputations: Reliability, validity, and responsiveness to change," *Archives of Physical Medicine and Rehabilitation*, vol. 85, no. 8, pp. 1339–1344, 2004.

[53] H. J. Day, "The assessment and description of amputee activity," *Prosthet Orthot Int*, vol. 5, no. 1, pp. 23–8, 1981.

[54] F. A. De Laat, L. D. Roorda, J. H. Geertzen, and C. Rommers, "Test–retest reliability of the special interest group on amputation medicine/dutch working group on amputations and prosthetics mobility scale, in persons wearing a prosthesis after a lower-limb amputation," *Disability and Rehabilitation*, vol. 42, no. 12, pp. 1762–1766, 2020.

[55] Physiopedia, "Locomotor capabilities index-5," 2022. https://www.physio-pedia.com/Locomotor_Capabilities_Index-5.

[56] B. G. Callaghan, S. Sockalingam, S. P. Treweek, and M. E. Condie, "A post-discharge functional outcome measure for lower limb amputees," *Prosthetics amp; Orthotics International*, vol. 26, no. 2, pp. 113–119, 2002.

[57] R. S. Gailey, K. E. Roach, E. B. Applegate, B. Cho, B. Cunniffe, S. Licht, M. Maguire, and M. S. Nash, "The amputee mobility predictor: An instrument to assess determinants of the lower-limb amputee's ability to ambulate," *Archives of Physical Medicine and Rehabilitation*, vol. 83, no. 5, pp. 613–627, 2002.

[58] A. Chadwell, L. Diment, M. Micó-Amigo, Z. M. R. Dafne, A. Dickinson, M. Granat, L. Kenney, K. Sisary, M. Sobuh, R. Ssekitoleko, and P. Worsley, "Technology for monitoring everyday prosthesis use: a systematic review," *Journal of Neuroengineering and Rehabilitation*, vol. 17, pp. 1–26, 2020.

[59] E. S. Arch, O. Erol, C. Bortz, C. Madden, M. Galbraith, A. Rossi, J. Lewis, J. S. Higginson, J. M. Buckley, and J. Horne, "Method to quantify cadence variability of individuals with lower-limb amputation," *JPO: Journal of Prosthetics and Orthotics*, vol. 29, no. 2, pp. 73–79, 2017.

[60] M. V. Albert, C. McCarthy, J. Valentin, M. Herrmann, K. Kording, and A. Jayaraman, "Monitoring functional capability of individuals with lower limb amputations using mobile phones," *PLoS One*, vol. 8, no. 6, 2013.

[61] M. V. Albert, S. Deeny, C. McCarthy, J. Valentin, and A. Jayaraman, "Monitoring daily function in persons with transfemoral amputations using a commercial activity monitor: a feasibility study," *Pm r*, vol. 6, no. 12, pp. 1120–7, 2014.

[62] B. Godfrey, J. Berdan, M. N. Kirk, and T. R. Chou, "The accuracy and validity of modus trex activity monitor in determining functional level in veterans with transtibial amputations," *JPO: Journal of Prosthetics and Orthotics*, vol. 30, no. 1, pp. 20–30, 2018.

[63] M. S. Orendurff, T. Kobayashi, C. Q. Villarosa, K. L. Coleman, and D. A. Boone, "Comparison of a computerized algorithm and prosthetists' judgment in rating functional levels based on daily step activity in transtibial amputees," *J Rehabil Assist Technol Eng*, vol. 3, 2016.

[64] M. S. Orendurff, S. U. Raschke, L. Winder, D. Moe, D. A. Boone, and T. Kobayashi, "Functional level assessment of individuals with transtibial limb loss: Evaluation in the clinical setting versus objective community ambulatory activity," *Journal of Rehabilitation and Assistive Technologies Engineering*, vol. 3, no. 0, 2016.

[65] D. W. Gardner, C. B. Redd, J. C. Cagle, B. J. Hafner, and J. E. Sanders, "Monitoring prosthesis user activity and doffing using an activity monitor and proximity sensors," *JPO: Journal of Prosthetics and Orthotics*, vol. 28, no. 2, pp. 68–77, 2016.

[66] N. Ramstrand and K. Nilsson, "Validation of a patient activity monitor to quantify ambulatory activity in an amputee population," *Prosthetics Orthotics International*, vol. 31, no. 2, pp. 157–166, 2007.

[67] M. T. M. Redfield, J. C. B. S. E. Cagle, B. J. P. Hafner, and J. E. P. Sanders, "Classifying prosthetic use via accelerometry in persons with transtibial amputations," *Journal of Rehabilitation Research and Development*, vol. 50, no. 9, pp. 1201–12, 2013.

[68] E. S. Arch, O. Erol, C. Bortz, C. Madden, M. Galbraith, A. Rossi, J. Lewis, J. S. Higginson, J. M. Buckley, and J. Horne, "Real-world walking performance of individuals with lower-limb amputation classified as medicare functional classification

level 2 and 3," *JPO: Journal of Prosthetics and Orthotics*, vol. 28, no. 2, pp. 51–57, 2016.

[69] A. Jayaraman, S. Deeny, Y. Eisenberg, G. Mathur, and T. Kuiken, "Global position sensing and step activity as outcome measures of community mobility and social interaction for an individual with a transfemoral amputation due to dysvascular disease," *Phys Ther*, vol. 94, no. 3, pp. 401–10, 2014.

[70] N. Shawen, L. Lonini, C. K. Mummidisetty, I. Shparii, M. V. Albert, K. Kording, and A. Jayaraman, "Fall detection in individuals with lower limb amputations using mobile phones: Machine learning enhances robustness for real-world applications," *JMIR mHealth and uHealth*, vol. 5, no. 10, 2017.

[71] L. Frossard, N. Stevenson, J. Smeathers, E. Häggström, K. Hagberg, J. Sullivan, D. Ewins, D. L. Gow, S. Gray, and R. Brånemark, "Monitoring of the load regime applied on the osseointegrated fixation of a trans-femoral amputee," *Prosthetics Orthotics International*, vol. 32, no. 1, pp. 68–78, 2008.

[72] A. Tura, M. Raggi, L. Rocchi, A. G. Cutti, and L. Chiari, "Gait symmetry and regularity in transfemoral amputees assessed by trunk accelerations," *Journal of NeuroEngineering and Rehabilitation*, vol. 7, no. 1, p. 4, 2010.

[73] J. Camargo, W. Flanagan, N. Csomay-Shanklin, B. Kanwar, and A. Young, "A machine learning strategy for locomotion classification and parameter estimation using fusion of wearable sensors," *IEEE Transactions on Biomedical Engineering*, vol. 68, no. 5, pp. 1569–1578, 2021.

[74] B. Chen, E. Zheng, Q. Wang, and L. Wang, "A new strategy for parameter optimization to improve phase-dependent locomotion mode recognition," *Neurocomputing*, vol. 149, pp. 585–593, 2015.

[75] P. S. Lum, L. Shu, E. M. Bochniewicz, T. Tran, L.-C. Chang, J. Barth, and A. W. Dromerick, "Improving accelerometry-based measurement of functional use of the upper extremity after stroke: Machine learning versus counts threshold method," *Neurorehabilitation and Neural Repair*, vol. 34, no. 12, pp. 1078–1087, 2020.

[76] A. A. Ibrahim, A. Küderle, H. Gaßner, J. Klucken, B. M. Eskofier, and F. Kluge, "Inertial sensor-based gait parameters reflect patient-reported fatigue in multiple sclerosis," *Journal of NeuroEngineering and Rehabilitation*, vol. 17, no. 1, 2020.

[77] A. Rampp, J. Barth, S. Schuelein, K.-G. Gassmann, J. Klucken, and B. M. Eskofier, "Inertial sensor-based stride parameter calculation from gait sequences in geriatric patients," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 4, pp. 1089–1097, 2015.

[78] J. Rudisch, T. Jöllenbeck, L. Vogt, T. Cordes, T. J. Klotzbier, O. Vogel, and B. Wollesen, "Agreement and consistency of five different clinical gait analysis systems in the assessment of spatiotemporal gait parameters," *Gait Posture*, vol. 85, pp. 55–64, 2021.

[79] Z. Q. Al-Busaidi, "Qualitative research and its uses in health care," *Sultan Qaboos Univ Med J*, vol. 8, no. 1, pp. 11–9, 2008.

[80] P. Glasziou, "Paul glasziou and iain chalmers: Is 85 percent of health research really "wasted"," *The BMJ*, 2016.

[81] P. Slattery, A. K. Saeri, and P. Bragge, "Research co-design in health: a rapid overview of reviews," *Health Research Policy and Systems*, vol. 18, no. 1, p. 17, 2020.

[82] L. M. Vaughn and F. Jacquez, "Participatory research methods – choice points in the research process," *Journal of Participatory Research Methods*, vol. 1, no. 1, 2020.

[83] C. Williamson, *Questionnaires, individual interviews and focus groups*, pp. 349–372. Tilde University Press, 2013.

[84] D. L. Morgan, R. A. Krueger, and J. A. King, *The focus group guidebook*. Sage, 1998.

[85] B. M. Wildemuth, *Applications of social research methods to questions in information and library science*. Bloomsbury Publishing USA, 2016.

[86] H. Kallio, A.-M. Pietilä, M. Johnson, and M. Kangasniemi, "Systematic methodological review: developing a framework for a qualitative semi-structured interview guide," *Journal of Advanced Nursing*, vol. 72, no. 12, pp. 2954–2965, 2016.

[87] V. Braun and V. Clarke, "Using thematic analysis in psychology," *Qualitative Research in Psychology*, vol. 3, no. 2, pp. 77–101, 2006.

[88] W. W. Hoeger, L. Bond, L. Ransdell, J. M. Shimon, and S. Merugu, "One-mile step count at walking and running speeds," *ACSM's Health Fitness Journal*, vol. 12, no. 1, pp. 14–19, 2008.

[89] D. R. Bassett, L. P. Toth, S. R. LaMunion, and S. E. Crouter, "Step counting: A review of measurement considerations and health-related applications," *Sports Medicine*, vol. 47, no. 7, pp. 1303–1315, 2017.

[90] S. A. Antos, M. K. Danilovich, A. R. Eisenstein, K. E. Gordon, and K. P. Kording, "Smartwatches can detect walker and cane use in older adults," *Innovation in Aging*, vol. 3, no. 1, 2019.

[91] T. Moder, K. Wisiol, and M. Wieser, "Walking aid identification using wearables," in *Computers Helping People with Special Needs*, Computers Helping People with Special Needs, pp. 335–341, Springer International Publishing, 2016.

[92] I. P. E. S. Putra and R. Vesilo, "Window-size impact on detection rate of wearable-sensor-based fall detection using supervised machine learning," in *2017 IEEE Life Sciences Conference (LSC)*, IEEE, 2017.

[93] M. Tazawa, H. Arii, Y. Ibe, H. Kobayashi, H. Chikuda, and N. Wada, "Effects of walking with a cane on frontal plane hip joint loading in patients with late-stage unilateral hip osteoarthritis," *Archives of Rehabilitation Research and Clinical Translation*, vol. 4, no. 3, p. 100209, 2022.

[94] Y. Zhang, C. Tao, X. Zhang, J. Guo, and Y. Fan, "Effects of cane use on the kinematic and kinetic of lower-extremity joints in inexperienced users," *Journal of Biomechanics*, vol. 146, p. 111426, 2023.

[95] B.-J. Hsue and F.-C. Su, "Gait and kinematics of the trunk and lower extremities in stair ascent using quadricane in healthy subjects," *Gait Posture*, vol. 29, no. 1, pp. 146–150, 2009.

[96] F. Labarrière, E. Thomas, L. Calistri, V. Optasanu, M. Gueugnon, P. Ornetti, and D. Laroche, "Machine learning approaches for activity recognition and/or activity prediction in locomotion assistive devices—a systematic review," *Sensors*, vol. 20, no. 21, p. 6345, 2020.

[97] V. Papapicco, B. Chen, M. Munih, A. Davalli, R. Sacchetti, E. Gruppioni, S. Crea, and N. Vitiello, "A classification approach based on directed acyclic graph to predict locomotion activities with one inertial sensor on the thigh," *IEEE Transactions on Medical Robotics and Bionics*, vol. 3, no. 2, pp. 436–445, 2021.

[98] D. Wang, X. Gu, and H. Yu, "A comparison of four neural networks algorithms on locomotion intention recognition of lower limb exoskeleton based on multi-source information," *Journal of Bionic Engineering*, vol. 21, no. 1, pp. 224–235, 2024.

[99] Z. Lu, A. Narayan, and H. Yu, "A deep learning based end-to-end locomotion mode detection method for lower limb wearable robot control," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 4091–4097, 2020.

[100] G. Khodabandelou, H. Moon, Y. Amirat, and S. Mohammed, "A fuzzy convolutional attention-based gru network for human activity recognition," *Engineering Applications of Artificial Intelligence*, vol. 118, p. 105702, 2023.

[101] J. Li, J. Zhang, K. Li, J. Cao, and H. Li, "A multimodal framework based on deep belief network for human locomotion intent prediction," *Biomedical Engineering Letters*, vol. 14, no. 3, pp. 559–569, 2024.

[102] Y. Gao, J. Zheng, L. Huang, H. Peng, L. Yin, Z. Yin, B. Wang, M. Peng, Z. Li, C. Zhang, and Y. Wang, "A novel bilstm-cnn method for pattern recognition in real time under triple physical loads in lower extremity exoskeleton," *IEEE Sensors Journal*, vol. 23, no. 14, pp. 15689–15701, 2023.

[103] C.-f. Chen, Z.-j. Du, L. He, Y.-j. Shi, J.-q. Wang, and W. Dong, "A novel gait pattern recognition method based on lstm-cnn for lower limb exoskeleton," *Journal of Bionic Engineering*, vol. 18, no. 5, pp. 1059–1072, 2021.

[104] L. Zhu, Z. Wang, Z. Ning, Y. Zhang, Y. Liu, W. Cao, X. Wu, and C. Chen, "A novel motion intention recognition approach for soft exoskeleton via imu," *Electronics*, vol. 9, no. 12, p. 2176, 2020.

[105] J. Tang, L. Zhao, M. Wu, Z. Jiang, J. Cao, and X. Bao, "A se-densenet-lstm model for locomotion mode recognition in lower limb exoskeleton," *PeerJ Comput Sci*, vol. 10, p. e1881, 2024.

[106] S. Gao, Y. Wang, C. Fang, and L. Xu, "A smart terrain identification technique based on electromyography, ground reaction force, and machine learning for lower limb rehabilitation," *Applied Sciences*, vol. 10, no. 8, p. 2638, 2020.

[107] E. Zheng, J. Wan, S. Gao, and Q. Wang, "Adaptive locomotion transition recognition with wearable sensors for lower limb robotic prosthesis," *IEEE/ASME Transactions on Mechatronics*, vol. 29, no. 1, pp. 279–289, 2024.

[108] S. Yu, J. Yang, T.-H. Huang, J. Zhu, C. J. Visco, F. Hameed, J. Stein, X. Zhou, and H. Su, "Artificial neural network-based activities classification, gait phase estimation, and prediction," *Annals of Biomedical Engineering*, vol. 51, no. 7, pp. 1471–1484, 2023.

[109] P. Chauhan, A. K. Singh, and N. K. Raghuwanshi, "Classifying the walking pattern of humans on different surfaces using convolutional features and shallow machine learning classifiers," *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, p. 09544062241247956, 2024. doi: 10.1177/09544062241247956.

[110] H. T. T. Vu, H. L. Cao, D. Dong, T. Verstraten, J. Geeroms, and B. Vanderborght, "Comparison of machine learning and deep learning-based methods for locomotion mode recognition using a single inertial measurement unit," *Front Neurorobot*, vol. 16, p. 923164, 2022.

[111] L. Sun and C. Li, "Convolutional neural network model for gait classification of flexible lower limb exoskeleton," in *2022 IEEE International Conference on Mechatronics and Automation (ICMA)*, pp. 1178–1182, 2022.

[112] L. Madaoui, O. Kerdjidj, and M. Kedir-Talha, "Design and implementation of imu-based locomotion mode recognition system on zynq soc," *Microprocessors and Microsystems*, vol. 102, p. 104927, 2023.

[113] Y. Han, C. Liu, L. Yan, and L. Ren, "Design of decision tree structure with improved bpnn nodes for high-accuracy locomotion mode recognition using a single imu," *Sensors (Basel)*, vol. 21, no. 2, 2021.

[114] H.-W. Oh and Y.-D. Hong, "Divergent component of motion-based gait intention detection method using motion information from single leg," *Journal of Intelligent Robotic Systems*, vol. 107, no. 4, p. 51, 2023.

[115] S. Luo, X. Shu, H. Zhu, and H. Yu, "Early prediction of lower limb prostheses locomotion mode transition based on terrain recognition," *IEEE Sensors Journal*, vol. 23, no. 22, pp. 27941–27948, 2023.

[116] B. Zhong, R. L. d. Silva, M. Tran, H. Huang, and E. Lobaton, "Efficient environmental context prediction for lower limb prostheses," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 52, no. 6, pp. 3980–3994, 2022.

[117] O. A. Gonzales-Huisa, G. Oshiro, V. E. Abarca, J. G. Chavez-Echajaya, and D. A. Elias, "Emg and imu data fusion for locomotion mode classification in transtibial amputees," *Prosthesis*, vol. 5, no. 4, pp. 1232–1256, 2023.

[118] S. Yin, X. Chen, T. Ma, Y. Wang, Y. Guo, Y. Leng, and C. Fu, "Environmental recognition and multimode continuous-phase control for a powered transfemoral prosthesis," in *2023 IEEE International Conference on Development and Learning (ICDL)*, pp. 55–60, 2023.

[119] Z. Qi, Q. Song, Y. Liu, and C. Guo, "Fsm-hsvm-based locomotion mode recognition for exoskeleton robot," *Applied Sciences*, vol. 12, no. 11, p. 5483, 2022.

[120] D. Marcos Mazon, M. Groefsema, L. R. B. Schomaker, and R. Carloni, "Imu-based classification of locomotion modes, transitions, and gait phases with convolutional recurrent neural networks," *Sensors*, vol. 22, no. 22, p. 8871, 2022.

[121] J. Bruinsma and R. Carloni, "Imu-based deep neural networks: Prediction of locomotor and transition intentions of an osseointegrated transfemoral amputee," *IEEE Trans Neural Syst Rehabil Eng*, vol. 29, pp. 1079–1088, 2021.

[122] F. Gao, G. Liu, F. Liang, and W. H. Liao, "Imu-based locomotion mode identification for transtibial prostheses, orthoses, and exoskeletons," *IEEE Trans Neural Syst Rehabil Eng*, vol. 28, no. 6, pp. 1334–1343, 2020.

[123] D. Shin, S. Lee, and S. Hwang, "Locomotion mode recognition algorithm based on gaussian mixture model using imu sensors," *Sensors*, vol. 21, no. 8, p. 2785, 2021.

[124] C.-S. Son and W.-S. Kang, "Multivariate cnn model for human locomotion activity recognition with a wearable exoskeleton robot," *Bioengineering*, vol. 10, no. 9, p. 1082, 2023.

[125] Y. Liu, H. An, H. Ma, and Q. Wei, "Novel feature extraction and locomotion mode classification using intelligent lower-limb prosthesis," *Machines*, vol. 11, no. 2, p. 235, 2023.

[126] M. R. Haque, M. H. Imtiaz, X. Shen, and E. Sazonov, "Real time level ground walking vs stair-climbing locomotion mode detection," in *2020 IEEE SENSORS*, pp. 1–4, 2020.

[127] X. Liu and Q. Wang, "Real-time locomotion mode recognition and assistive torque control for unilateral knee exoskeleton on different terrains," *IEEE/ASME Transactions on Mechatronics*, vol. 25, no. 6, pp. 2722–2732, 2020.

[128] R. Guo, W. Li, Y. He, T. Zeng, B. Li, G. Song, and J. Qiu, "Terrain slope parameter recognition for exoskeleton robot in urban multi-terrain environments," *Complex Intelligent Systems*, vol. 10, no. 2, pp. 3107–3118, 2024.

[129] Y. Qian, C. Chen, J. Xiong, Y. Wang, Y. Leng, H. Yu, and C. Fu, "Terrain-adaptive exoskeleton control with predictive gait mode recognition: A pilot study during level walking and stair ascent," *IEEE Transactions on Medical Robotics and Bionics*, vol. 6, no. 1, pp. 281–291, 2024.

[130] L. Du, F. Zhang, M. Liu, and H. Huang, "Toward design of an environment-aware adaptive locomotion-mode-recognition system," *IEEE Trans Biomed Eng*, vol. 59, no. 10, pp. 2716–25, 2012.

[131] T. G. Dietterich, "Machine learning for sequential data: A review," in *Structural, Syntactic, and Statistical Pattern Recognition*, Structural, Syntactic, and Statistical Pattern Recognition, pp. 15–30, Springer Berlin Heidelberg, 2002.

[132] K. S. Türker, "Electromyography: Some methodological problems and issues," *Physical Therapy*, vol. 73, no. 10, pp. 698–710, 1993.

[133] A. Saidi, S. Ben Othman, M. Dhouibi, and S. Ben Saoud, "Fpga-based implementation of classification techniques: A survey," *Integration*, vol. 81, pp. 280–299, 2021.

[134] R. Choudhary and H. K. Gianey, "Comprehensive review on supervised machine learning algorithms," in *2017 International Conference on Machine Learning and Data Science (MLDS)*, pp. 37–43, 2017.

[135] IBM, "What is linear discriminant analysis (lda)?," 27 November 2023 2023. https://www.ibm.com/topics/linear-discriminant-analysis#:~:text=Linear%20discriminant%20analysis%20(LDA)%20is,helps%20optimize%20machine%20learning%20models.

[136] IBM, "What are support vector machines (svms)?," 27 December 2023 2023. https://www.ibm.com/topics/support-vector-machine#:~:text=What%20are%20SVMs%3F,in%20an%20N%2Ddimensional%20space.

[137] U. of Cincinnati-Business-Analytics, "Linear quadratic discriminant analysis," 2024. https://uc-r.github.io/discriminant_analysis.

[138] L. E. Chai, S. K. Loh, S. T. Low, M. S. Mohamad, S. Deris, and Z. Zakaria, "A review on the computational approaches for gene regulatory network construction," *Computers in Biology and Medicine*, vol. 48, pp. 55–65, 2014.

[139] One-Off-Coder, "11. dynamic bayesian networks, hidden markov models," Apr 03, 2024 2024. https://datascience.oneoffcoder.com/dbn-hmm.html.

[140] IBM, "What is a decision tree?," 15th August 2024 2024. https://www.ibm.com/topics/decision-trees.

[141] M. Belgiu and L. Drăguţ, "Random forest in remote sensing: A review of applications and future directions," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 114, pp. 24–31, 2016.

[142] S. Walczak and N. Cerpa, *Artificial Neural Networks*, pp. 631–645. New York: Academic Press, 2003.

[143] IBM, "What is a neural network?," 2nd October 2024 2024. https://www.ibm.com/topics/neural-networks.

[144] A. Das, P. K. Kumawat, and N. D. Chaturvedi, *A Study to Target Energy Consumption in Wastewater Treatment Plant using Machine Learning Algorithms*, vol. 50, pp. 1511–1516. Elsevier, 2021.

[145] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: an overview and application in radiology," *Insights into Imaging*, vol. 9, no. 4, pp. 611–629, 2018.

[146] S. Kannan, K. Subbaram, and M. Faiyazuddin, *Chapter 17 - Artificial intelligence in vaccine development: Significance and challenges ahead*, pp. 467–486. Academic Press, 2023.

[147] J. A. Spanias, A. M. Simon, K. A. Ingraham, and L. J. Hargrove, "Effect of additional mechanical sensor data on an emg-based pattern recognition system for a powered leg prosthesis," in *2015 7th International IEEE/EMBS Conference on Neural Engineering (NER)*, pp. 639–642, 2015.

[148] J. A. Spanias, E. J. Perreault, and L. J. Hargrove, "Detection of and compensation for emg disturbances for powered lower limb prosthesis control," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 24, no. 2, pp. 226–234, 2016.

[149] J. A. Spanias, A. M. Simon, E. J. Perreault, and L. J. Hargrove, "Preliminary results for an adaptive pattern recognition system for novel users using a powered lower limb prosthesis," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 5083–5086, 2016.

[150] J. A. Spanias, A. M. Simon, and L. J. Hargrove, "Across-user adaptation for a powered lower limb prosthesis," in *IEEE International Conference on Rehabilitation Robotics (ICORR)*, IEEE, 2017.

[151] J. A. Spanias, A. M. Simon, S. B. Finucane, E. J. Perreault, and L. J. Hargrove, "Online adaptive neural control of a robotic lower limb prosthesis," *Journal of Neural Engineering*, vol. 15, p. 016015, jan 2018.

[152] A. J. Young, A. Simon, and L. J. Hargrove, "An intent recognition strategy for transfemoral amputee ambulation across different locomotion modes," in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 1587–1590, 2013.

[153] A. J. Young, A. M. Simon, N. P. Fey, and L. J. Hargrove, "Intent recognition in a powered lower limb prosthesis using time history information," *Annals of biomedical engineering*, vol. 42, pp. 631–641, 2014.

[154] A. J. Young, T. A. Kuiken, and L. J. Hargrove, "Analysis of using emg and mechanical sensors to enhance intent recognition in powered lower limb prostheses," *Journal of Neural Engineering*, vol. 11, p. 056021, sep 2014.

[155] A. J. Young and L. J. Hargrove, "A classification method for user-independent intent recognition for transfemoral amputees using powered lower limb prostheses," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 24, no. 2, pp. 217–225, 2016.

[156] K.-H. Chen, P.-C. Chen, K.-C. Liu, and C.-T. Chan, "Wearable sensor-based rehabilitation exercise assessment for knee osteoarthritis," *Sensors*, vol. 15, no. 2, pp. 4193–4211, 2015.

[157] A. Jamieson, L. Murray, L. Stankovic, V. Stankovic, and A. Buis, "Human activity recognition of individuals with lower limb amputation in free-living conditions: A pilot study," *Sensors (Basel)*, vol. 21, no. 24, 2021.

[158] B.-Y. Su, J. Wang, S.-Q. Liu, M. Sheng, J. Jiang, and K. Xiang, "A cnn-based method for intent recognition using inertial measurement units and intelligent lower limb prosthesis," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 5, pp. 1032–1042, 2019.

[159] Y. Luo, S. M. Coppola, P. C. Dixon, S. Li, J. T. Dennerlein, and B. Hu, "A database of human gait performance on irregular and uneven surfaces collected by wearable sensors," *Scientific Data*, vol. 7, no. 1, 2020.

[160] A. Logacjov, K. Bach, A. Kongsvold, H. B. Bårdstu, and P. J. Mork, "Harth: A human activity recognition dataset for machine learning," *Sensors*, vol. 21, no. 23, p. 7853, 2021.

[161] A. Mai and S. Commuri, "Gait identification for an intelligent prosthetic foot," in *IEEE International Symposium on Intelligent Control (ISIC)*, IEEE, 2011.

[162] Z. Fan and H. He, "Source selection for real-time user intent recognition toward volitional control of artificial legs," *IEEE Journal of Biomedical and Health Informatics*, vol. 17, no. 5, pp. 907–914, 2013.

[163] M. Mendez, "Advantages and disadvantages of inertial measurement units," 15/9/2023 2023.

[164] C. E. Shell, G. K. Klute, and R. R. Neptune, "Identifying classifier input signals to predict a cross-slope during transtibial amputee walking," *PLOS ONE*, vol. 13, no. e0192950, 2018.

[165] J. Mai, D. Xu, H. Li, S. Zhang, J. Tan, and Q. Wang, "Implementing a soc-fpga based acceleration system for on-board svm training for robotic transtibial prostheses," in *IEEE International Conference on Real-time Computing and Robotics (RCAR)*, IEEE, 2018.

[166] F. Sherratt, A. Plummer, and P. Iravani, "Understanding lstm network behaviour of imu-based locomotion mode recognition for applications in prostheses and wearables," *Sensors*, vol. 21, no. 4, p. 1264, 2021.

[167] B. Hu, S. Li, Y. Chen, R. Kavi, and S. Coppola, "Applying deep neural networks and inertial measurement unit in recognizing irregular walking differences in the real world," *Applied Ergonomics*, vol. 96, p. 103414, 2021.

[168] P. Dixon and V. Shah, "An automatic approach for classifying gait on outdoor walking surfaces based on wearable sensors," *Gait Posture*, vol. 90, pp. 46–47, 2021.

[169] B. Griffiths, L. Diment, and M. H. Granat, "A machine learning classification model for monitoring the daily physical behaviour of lower-limb amputees," *Sensors*, vol. 21, no. 22, p. 7458, 2021.

[170] X. Wang, Q. Wang, E. Zheng, K. Wei, and L. Wang, *A Wearable Plantar Pressure Measurement System: Design Specifications and First Experiments with an Amputee*, pp. 273–281. Springer Berlin Heidelberg, 2013.

[171] B. Chen, X. Wang, Y. Huang, K. Wei, and Q. Wang, "A foot-wearable interface for locomotion mode recognition based on discrete contact force distribution," *Mechatronics*, vol. 32, pp. 12–21, 2015.

[172] Z. Liu, W. Lin, Y. Geng, and P. Yang, "Intent pattern recognition of lower-limb motion based on mechanical sensors," *IEEE/CAA Journal of Automatica Sinica*, vol. 4, no. 4, pp. 651–660, 2017.

[173] D. Xu, Y. Feng, J. Mai, and Q. Wang, "Real-time on-board recognition of continuous locomotion modes for amputees with robotic transtibial prostheses," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 10, pp. 2015–2025, 2018.

[174] N. Constant, G. Cay, V. Ravichandran, R. Diouf, U. Akbar, and K. Mankodiya, *Chapter 13 - Data analytics for wearable IoT-based telemedicine*, pp. 357–378. Academic Press, 2021.

[175] R. Pascanu, "How to construct deep recurrent neural networks," *arXiv preprint arXiv:1312.6026*, 2013.

[176] HM-Government, "Approved document k: Protection from falling, collision and impact," 2010. https://www.gov.uk/government/publications/protection-from-falling-collision-and-impact-approved-document-k.

[177] N. Mohapatra, K. Shreya, and A. Chinmay, "Optimization of the random forest algorithm," in *Advances in Data Science and Management* (S. Borah, V. Emilia Balas, and Z. Polkowski, eds.), pp. 201–208, Springer Singapore, 2020.

[178] S. B. Thies, S. Bevan, M. Wassall, B. K. Shajan, L. Chowalloor, L. Kenney, and D. Howard, "Evaluation of a novel biomechanics-informed walking frame, developed through a knowledge transfer partnership between biomechanists and design engineers," *BMC Geriatrics*, vol. 23, no. 1, p. 734, 2023.

[179] J. W. Youdas, B. J. Kotajarvi, D. J. Padgett, and K. R. Kaufman, "Partial weight-bearing gait using conventional assistive devices," *Archives of Physical Medicine and Rehabilitation*, vol. 86, no. 3, pp. 394–398, 2005. doi: 10.1016/j.apmr.2004.03.026.

[180] F. El Barakaz, O. Boutkhoum, and A. El Moutaouakkil, "Feature selection method based on classification performance score and p-value," in *Advanced Intelligent Systems for Sustainable Development (AI2SD'2020)*, pp. 379–386, Springer International Publishing, 2022.

[181] K. Y. Tong and M. H. Granat, "Virtual artificial sensor technique for functional electrical stimulation," *Med Eng Phys*, vol. 20, no. 6, pp. 458–68, 1998.

[182] P. A. Tresadern, S. B. Thies, L. P. Kenney, D. Howard, C. Smith, J. Rigby, and J. Y. Goulermas, "Simulating acceleration from stereophotogrammetry for medical device design," *J Biomech Eng*, vol. 131, no. 6, p. 061002, 2009.

[183] M. C. Ruder, M. A. Hunt, J. M. Charlton, C. T. F. Tse, and D. Kobsar, "Original article: Validity and reliability of gait metrics derived from researcher-placed and self-placed wearable inertial sensors," *Journal of Biomechanics*, vol. 142, p. 111263, 2022.

[184] T. Tan, D. P. Chiasson, H. Hu, and P. B. Shull, "Influence of imu position and orientation placement errors on ground reaction force estimation," *Journal of Biomechanics*, vol. 97, p. 109416, 2019.

[185] C. Lützner, H. Voigt, I. Roeder, S. Kirschner, and J. Lützner, "Placement makes a difference: Accuracy of an accelerometer in measuring step number and stair climbing," *Gait Posture*, vol. 39, no. 4, pp. 1126–1132, 2014.

[186] T. V. Barreira, D. A. Rowe, and M. Kang, "Parameters of walking and jogging in healthy young adults," *International Journal of Exercise Science*, vol. 3, no. 1, pp. 4–13, 2010.

[187] B. Langley, A. Jones, T. Board, and M. Greig, "Modified conventional gait model vs. six degrees of freedom model: A comparison of lower limb kinematics and associated error," *Gait Posture*, vol. 89, pp. 1–6, 2021.

[188] C. Ding and H. Peng, "Minimum redundancy feature selection from microarray gene expression data," *J Bioinform Comput Biol*, vol. 3, no. 2, pp. 185–205, 2005.

[189] MathWorks, "Rank features for classification using minimum redundancy maximum relevance (mrmr) algorithm," 2024. https://uk.mathworks.com/help/stats/fscmrmr.html.

[190] J. Goldberger, G. E. Hinton, S. R. Roweis, and R. Salakhutdinov, "Neighbourhood components analysis," *Advances in Neural Information Processing Systems 17 (NIPS 2004)*, 2004.

[191] MathWorks, "Feature selection using neighborhood component analysis for regression," 2024. https://uk.mathworks.com/help/stats/fsrnca.html.

[192] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Machine Learning*, vol. 46, no. 1, pp. 389–422, 2002.

[193] F. Iglesias and W. Kastner, "Analysis of similarity measures in times series clustering for the discovery of building energy patterns," *Energies*, vol. 6, no. 2, pp. 579–597, 2013.

[194] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*. Taylor Francis, 2013.

[195] D. Loudon, A. Taylor, and A. S. Macdonald, *The Use of Qualitative Design Methods in the Design, Development and Evaluation of Virtual Technologies for Healthcare: Stroke Case Study*, pp. 371–390. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014.

[196] L. F. Teixeira-Salmela, S. Nadeau, M.-H. Milot, D. Gravel, and L. F. Requião, "Effects of cadence on energy generation and absorption at lower extremity joints during gait," *Clinical Biomechanics*, vol. 23, no. 6, pp. 769–778, 2008.

[197] C. De Marchis, S. Ranaldi, T. Varrecchia, M. Serrao, S. F. Castiglia, A. Tatarelli, A. Ranavolo, F. Draicchio, F. Lacquaniti, and S. Conforto, "Characterizing the gait of people with different types of amputation and prosthetic components through multimodal measurements: A methodological perspective," *Front Rehabil Sci*, vol. 3, p. 804746, 2022.

[198] P. Tamburini, F. Storm, C. Buckley, M. C. Bisi, R. Stagni, and C. Mazzà, "Moving from laboratory to real life conditions: Influence on the assessment of variability and stability of gait," *Gait Posture*, vol. 59, pp. 248–252, 2018.

[199] A. S. McIntosh, K. T. Beatty, L. N. Dwan, and D. R. Vickers, "Gait dynamics on an inclined walkway," *Journal of Biomechanics*, vol. 39, no. 13, pp. 2491–2502, 2006.

[200] G. Strutzenberger, L. Leutgeb, L. Claußen, and H. Schwameder, "Gait on slopes: Differences in temporo-spatial, kinematic and kinetic gait parameters between walking on a ramp and on a treadmill," *Gait Posture*, vol. 91, pp. 73–78, 2022.

[201] N. D. A. Thomas, J. D. Gardiner, R. H. Crompton, and R. Lawson, "Physical and perceptual measures of walking surface complexity strongly predict gait and gaze behaviour," *Human Movement Science*, vol. 71, p. 102615, 2020.

Appendix A

Outcome measures from Balk et al

A.1 Outcome measures from Balk et al

Outcome measures that Balk et al. investigated:

- 1 Leg Standing Balance
- 180 Degree Turn Test
- 2MWT (2 Minute Walk Test)
- 6MWT (6 Minute Walk Test)
- AAS (Amputee Activity Survey)
- ABC (Activities-specific Balance Confidence)
- ADAPT (Assessment of Daily Activity Performance in Transfemoral Amputees)
- AMP (Amputee Mobility Predictor with, AMPPRO, or without prosthesis, AMPnoPRO)
- AMPSIMM (Amputee Single Item Mobility Measure)
- Barthel Index
- BBS (Berg Balance Scale)
- Climbing Stairs Questionnaire ES-3
- Employment Questionnaire
- FAC (Functional Ambulation Categories)

- FAI (Frenchay Activities Index)
- FIM (Functional Independence Measure)
- FSST (Four Square Step Test)
- Functional Reach Test
- Houghton Scale
- L Test (L Test of Functional Mobility)
- LCI (Locomotor Capabilities Index)
- LEMOCOT (Lower-Extremity Motor Coordination Test)
- NQ-ACGC (Quality of Life in Neurological Conditions – Applied Cognition/General Concerns)
- OPCS (Office of Population Censuses and Surveys Scale)
- OPUS (Orthotics Prosthetics Users Survey)
- Patient Activity Monitor
- PEQ, PEQ-MS (Prosthetic Evaluation Questionnaire, Mobility Subscale)
- PFI (Physical Function Index)
- PGI (Patient Generated Index)
- PLUS-M (Prosthetic Limb Users Survey of Mobility)
- PPA (Prosthetic Profile of the Amputee)
- PROMIS-29 (Patient-Reported Outcomes Measurement Information System 29-Item Profile)
- PROS (Prosthetist's Perception of Client's Ambulatory Abilities)
- PSFS (Patient-Specific Functional Scale)
- Q-TFA (Questionnaire for Persons with a Transfemoral Amputation)
- Rising and Sitting Down Questionnaire
- RMI (Rivermead Mobility Index)

- SAT-PRO (Satisfaction with Prosthesis Questionnaire)
- SCS (Socket Comfort Score)
- SF-12/SF-36/SF-36V (Short Form Health Surveys 12, 36, and 36V)
- SIGAM (Special Interest Group of Amputation Medicine)
- Single beam test
- SIP-PD (Sickness Impact Profile-Physical Dimension)
- Tandem Test
- TAPES (Trinity Amputation and Prosthesis Experience Scales)
- TFP (Transfemoral Fitting Predictor)
- TUG (Timed Up and Go)
- TWT (Timed Walking Test)
- Walking Questionnaire
- WHOQOL-BREF (World Health Organization Quality-of-Life Scale – Brief Version)

Appendix B

Components Limbs 4 Life recommend for each K level

B.1 Components Limbs 4 Life recommend for each K level

Table B.1 Components Limbs 4 Life recommend for each K level [8]

K1	K2	K3	K4
Massons TeuFel WLD31	OPC Knee Legworks All Terrain Premium	Ossur POWERKNEE Microprocessor Knee	OttoBock Genuim Knee Microprocessor
OPC Knee Proteor 1M112	OPC Proteor Matik Pneumatic	OPC Pliè3 Knee	OttoBock 3R80 Knee Hydraulic
Ossur Locking Knee Mech	Ossur Balance Knee OFM1	Ossur Knee RHEO XC Microprocessor	OPC BioDaptinc Moto Knee Sports
OttoBock 3 R31 Prosedo Knee hydraulic	Ossur Balance Knee OFM2	OttoBock 3E80 Knee MicroProcessor	OPC Proteor EasyRide Knee Sports
OttoBock 3 R41 Knee Mechanical	Ossur Knee OP4 Pneumatic	OttoBock 3R60 Knee Pneumatic	Ossur Cheetah Knee Hydraulic
OttoBock 1s101 SACH	Ossur Total Knee 1900	OttoBock 3 R78 Knee Pneumatic	OttoBock 3S80 KneeSports
OttoBock 1M10 Adjust Articulated	OttoBock Pheon 3 R62 Knee mechanical	OttoBock 3R92 Knee Pneumatic	OttoBock Pro Carve KneeSports

OPC Seattle natural SACH	OttoBock 3 R90 Knee Mechanical	OttoBock C Leg4 Knee MicroProcessor	Massons WillowWood MetaArc Dynamic
Ossur Flex-Foot Balance K1 articulated	OttoBock Kenevo Knee MicroProcessor	OPC Blatchford Orion3 MicroProcessor	Massons WillowWood MetaShock Dynamic
Massons WillowWood SACH foot	Massons WillowWood DuraWalk multi axis	OPC CollegePark Capital Knee Hydraulic	OPC Fillauer aeris performance 2 Dynamic
Massons WillowWood Single Axis foot	OPC Blatchford Multiflex	OPC Nabtesco Allux 2 Microprocessor	OPC Fillauer Allpro Dynamic
Ossur Flex foot balance with D/P flexion articulated	OPC College Park Breeze Foot SACH flexible	OPC Nabtesco Symphony Knee Hydraulic	OPC Rush Foot HiPro flex
	OPC college Park Odyssey Hydraulic	OPC Proteor HyTrek Hydraulic	Ossur Pro Flex XC
	OPC College Park Tribute multi axis	OPC Freedom Quattro Microprocessor	Ossur Pro Flex XC torsion Dynamic
	OPC Fillauer Foot	Ossur Mauch Knee Hydraulic	Ossur Reflex Shock Dynamic
	Ossur Balance foot J articulated	Ossur Paso Knee Pneumatic	OttoBock 1A1-1 Empower Microprocessor
	Ossur Balance Foot S torsion dynamic response	Ossur Total Knee 2000 Hydraulic	OttoBock IC50 Taleo Dynamic
	OttoBock IC11 Terion Dynamic	QLD Pros VGK Go Knee Fluidic	OttoBock IC61 Vertical Shock Dynamic
	OttoBock VS4 Kintrol Articulated	QLD Pros VKG S fluidic knee	OttoBock IC63 Triton LP Dynamic
	OttoBock VS5 Restore dynamic	QLDPros VGK X Fluidic lightweight	OttoBock Maverick Xtreme AT Dynamic

		Massons Willow-Wood Impulse Response foot	
		Massons WillowWood Koa Dynamic	
		OPC Freedom Innovation Kinterra Hydraulic	
		Ossur Proprio Microprocessor	
		Ossur Pro Flex Pivot Dynamic response	
		Ossur Talux Dynamic Response	
		OttoBock 1B1 Meridium Micrprocessor foot	
		OttoBock IC30 Trias Dynamic	
		OttoBock IC40 CWalk Dynamic	
		QLDPros Xtend Foot Dynamic	

Appendix C

interview guides

C.1 Interview guide 1

A study to understand perspectives of clinicians regarding activity levels and activity level assessments of lower limb amputees.

Interview Guide

Version (10-03-21)

Interviews are planned with 6 healthcare/clinical professionals who work with lower limb prosthetic patients, to gain an understanding of their experiences with activity level assessments. In particular, we will explore the way they conduct the assessments and how they distinguish between K2 and K3 activity levels. The interviews will also explore how the clinicians view activity level assessments, how activity levels change over time, and how the assessments could be improved. The intent of the interviews is to inform a questionnaire that will be distributed to gain further understanding of the issues.

Inclusion criteria: 1) conduct activity level assessments for lower limb prosthetic patients, 2) able to provide informed consent, 3) able to do an interview over the phone in the English language. *Exclusion criteria:* 1) the study finishes while they are deciding to take part.

A general opening question may be "What is your general experience with assessment of activity levels?", to be followed by a series of trigger questions, for example:

- What standards do you use to assess activity levels?
- What information or data do you use to decide on the patient K level?
- How do you distinguish between K2 and K3?
- Do you see activity levels change over time?
- Do you ever change a patient's K level?
- What information or data would be useful for better K level assignment?
- What activities do K2/3 patients do?
- Do you think there is anything a patient would not want to have measured?
- How would you improve activity assessments?
- What do you think a perfect activity assessment tool needs to take into account?
- How do kinematic variables (like stride length, cadence, stride symmetry, ext.) influence activity level assessments?
- How do you assess kinematic variables?
- How do kinetic variables (like ground reaction force, muscle activation, ext.) influence activity level assessments?
- How do you assess kinetic variables?
- How does activity of daily living affect activity level assessments?
- How do you assess activity of daily living?

Follow up questions may be asked such as:

- What do you think would be the consequences of this?
- How does this affect the patient/yourself?
- How do you feel about this?
- Does this worry you?
- What do you think may help with this problem?
- Why do you think this is?
- What are your views on this?

C.2 Interview guide 2

Interview Guide
Version (14-11-23)

Interviews are planned with all clinicians who assessed patients that took part in the real-world study to monitor lower limb prosthetic patient's activity levels using the data from the sensor system. These interviews will be conducted to explore the clinicians' experiences with using the sensor system.

The Inclusion Criteria: 1) assessed patients' K levels using the data from the sensor system

A general opening question may be "What are your thoughts on the sensor data?" which will be followed by a series of trigger questions throughout the interview, for example:

- Could you see any problems with the sensor system?
- Would the sensors affect your assessments?
- Do you find the data interesting?
- Would the data affect your activity level assessments?
- Does the data surprise you?
- Is the data useful?
- Is there any other data you would like?
- What are your thoughts on the presentation of the data?
- Is there any parts of the data you do not think are relevant?
- Are there any other improvement you would like?

For all interviewees, follow up questions may be asked such as:

- How did this affect you?
- How do you feel about this?
- Does this worry you?
- What do you think may help with this problem?
- Why do you think this is?
- What are your views on this?

Appendix D

Ethical approvals

D.1 University of Salford ethical approval 1

From: ethics <ethics@salford.ac.uk> Sent: 07 May 2021 10:05 To: Sibylle Thies <S.Thies@salford.ac.uk> Subject: Ethics Application: Panel Decision Importance: Low
The Ethics Panel has reviewed your application: A study to understand perspectives of clinicians regarding activity levels and activity level assessments of lower limb amputees.
Application ID: 1710
The decision is: Application Approved.
If the Chair has provided comments, these are as follows:
Please use the Ethics Application Tool to review your application.

D.2 University of Salford ethical approval 2

From: ethics Sent: 18 March 2022 10:35 To: Matthew Wassall Cc: Sibylle Thies Subject: App Ref. 5924: Ethics Application: Approval
Importance: Low
The Ethics Panel has reviewed your application: To develop a system to better understand the activities of lower limb prosthesis users in everyday life. Application ID: 4743
The decision is: Application Approved.
If the Chair has provided comments, these are as follows: N/A
You will no longer be able to edit your application in the system.

D.3 IRAS ethical approval

Ymchwil Iechyd a Gofal Cymru
Health and Care Research Wales

NHS
Health Research Authority

Mr Matthew Wassall
PhD Student
University of Salford
5 Hulton Street
Statham St
Salford
M5 3GE

Email: approvals@hra.nhs.uk
HCRW.approvals@wales.nhs.uk

14 July 2022

Dear Mr Wassall

HRA and Health and Care Research Wales (HCRW) Approval Letter

Study title:	**To develop a system to better understand the activities of lower limb prosthesis users in everyday life**
IRAS project ID:	**314743**
Protocol number:	**N/A**
REC reference:	**22/EM/0134**
Sponsor	**University of Salford**

I am pleased to confirm that **HRA and Health and Care Research Wales (HCRW) Approval** has been given for the above referenced study, on the basis described in the application form, protocol, supporting documentation and any clarifications received. You should not expect to receive anything further relating to this application.

Please now work with participating NHS organisations to confirm capacity and capability, in line with the instructions provided in the "Information to support study set up" section towards the end of this letter.

How should I work with participating NHS/HSC organisations in Northern Ireland and Scotland?
HRA and HCRW Approval does not apply to NHS/HSC organisations within Northern Ireland and Scotland.

If you indicated in your IRAS form that you do have participating organisations in either of these devolved administrations, the final document set and the study wide governance report (including this letter) have been sent to the coordinating centre of each participating nation. The relevant national coordinating function/s will contact you as appropriate.

Please see IRAS Help for information on working with NHS/HSC organisations in Northern Ireland and Scotland.

How should I work with participating non-NHS organisations?
HRA and HCRW Approval does not apply to non-NHS organisations. You should work with your non-NHS organisations to obtain local agreement in accordance with their procedures.

What are my notification responsibilities during the study?

The standard conditions document "*After Ethical Review – guidance for sponsors and investigators*", issued with your REC favourable opinion, gives detailed guidance on reporting expectations for studies, including:

- Registration of research
- Notifying amendments
- Notifying the end of the study

The HRA website also provides guidance on these topics, and is updated in the light of changes in reporting expectations or procedures.

Who should I contact for further information?
Please do not hesitate to contact me for assistance with this application. My contact details are below.

Your IRAS project ID is **314743**. Please quote this on all correspondence.

Yours sincerely,
Kelly Rowe

Approvals Manager

Email: approvals@hra.nhs.uk

Copy to: *Dr Sibylle Thies*

List of Documents

The final document set assessed and approved by HRA and HCRW Approval is listed below.

Document	*Version*	*Date*
Contract/Study Agreement template [Model Non commercial PIC agreement]		
Copies of materials calling attention of potential participants to the research [Flyer and poster]	1	01 June 2022
Evidence of Sponsor insurance or indemnity (non NHS Sponsors only) [Insurance certificate]	1	01 August 2021
IRAS Application Form [IRAS_Form_26052022]		26 May 2022
IRAS Application Form XML file [IRAS_Form_26052022]		26 May 2022
IRAS Checklist XML [Checklist_01062022]		01 June 2022
Other [Cover letter for IRAS 314743 approval response]	1	12 July 2022
Other [PIC agreement]	1	16 May 2022
Other [Local Covid risk assessment]	1	21 January 2022
Other [Data protection checklist]	1	21 January 2022
Other [Risk assessment]	1	05 May 2022
Participant consent form [Consent form]	2	07 March 2022
Participant information sheet (PIS) [PIS]	3	11 July 2022
Research protocol or project proposal [Study protocol]	2	12 July 2022
Summary CV for Chief Investigator (CI) [CV Matthew Wassall]		13 May 2022
Summary CV for supervisor (student research) [CV Malcolm Granat]	1	01 June 2022
Summary CV for supervisor (student research) [CV Sibylle Thies]	1	01 April 2022

IRAS project ID	314743

Information to support study set up

The below provides all parties with information to support the arranging and confirming of capacity and capability with participating NHS organisations in England and Wales. This is intended to be an accurate reflection of the study at the time of issue of this letter.

Types of participating NHS organisation	Expectations related to confirmation of capacity and capability	Agreement to be used	Funding arrangements	Oversight expectations	HR Good Practice Resource Pack expectations
Participating NHS organisations will be Participant Identification centres and provide information to patients.	Research activities should not commence at participating NHS organisations in England or Wales prior to their formal confirmation of capacity and capability to deliver the study in accordance with the contracting expectations detailed. Due to the nature of the activities involved, organisations will be expected to provide that confirmation to the sponsor Within **35**of receipt of the local information pack After HRA/HCRW	The sponsor has provided the appropriate model non-commercial PIC agreement that it intends to use as a contract between participating organisations and NHS organisations acting as their Participant Identification Centres (PICs).	No application for external funding has been made.	Neither a PI or local collaborator is expected at participating NHS organisations	The study is limited to NHS organisations acting as PICs and HR good practice arrangements are not expected for the trial.

	Approval has been issued. If the organisation is not able to formally confirm capacity and capability within this timeframe, they must inform the sponsor of this and provide a justification. If the sponsor is not satisfied with the justification, then the sponsor may escalate to the National Coordinating Function where the participating NHS organisation is located.				

Other information to aid study set-up and delivery

This details any other information that may be helpful to sponsors and participating NHS organisations in England and Wales in study set-up.
The applicant has indicated that they do not intend to apply for inclusion on the NIHR CRN Portfolio.

Appendix E

Variance Assessments results Terrain Classification

E.1 Variance Assessments results Terrain Classification

Table E.1 Results at lowest p values ankle. F=Flat ground, Su=Stair ascent, Sd=Stair descent, Ru= Ramp ascent, Rd=Ramp descent, Un=Uneven terrain

Terrain	F to Su	F to Sd	F to Ru	F to Rd	F to Un	Su to Sd	Su to Ru	Su to Rd	Su to Un	Sd to Ru	Sd to Rd	Sd to Un	Ru to Rd	Ru to Un	Rd to Un
% of stride	22	22	20	20	18	85	31	11	21	99	13	84	51	64	74
Feature	Free resultant	NPP Free resultant	Free X	NPP resultant	NPP X	NPP resultant	Free X	NPP X	NPP Free resultant	NPP Free Z	NPP X	NPP resultant	Free resultant	X	Z
SMD	2.10	2.52	0.52	0.93	0.61	0.79	1.42	1.25	1.05	1.36	1.55	1.06	0.61	0.55	0.50

Table E.2 Results at highest SMD ankle. F=Flat ground, Su=Stair ascent, Sd=Stair descent, Ru= Ramp ascent, Rd=Ramp descent, Un=Uneven terrain

Terrain	F to Su	F to Sd	F to Ru	F to Rd	F to Un	Su to Sd	Su to Ru	Su to Rd	Su to Un	Sd to Ru	Sd to Rd	Sd to Un	Ru to Rd	Ru to Un	Rd to Un
SMD	2.27	2.52	0.72	0.93	0.68	0.79	1.86	1.32	1.05	2.05	1.58	1.16	0.61	0.65	0.52
% of stride	22	22	20	20	20	85	21	34	21	14	12	86	51	10	77
Feature	NPP Free resultant	NPP Free resultant	NPP resultant	NPP resultant	NPP resultant	NPP resultant	NPP Free Z	NPP X	NPP Free resultant	NPP X	NPP X	NPP resultant	Free resultant	NPP X	Z

Table E.3 Results at lowest p values Mid shank. F=Flat ground, Su=Stair ascent, Sd=Stair descent, Ru= Ramp ascent, Rd=Ramp descent, Un=Uneven terrain

Terrain	F to Su	F to Sd	F to Ru	F to Rd	F to Un	Su to Sd	Su to Ru	Su to Rd	Su to Un	Sd to Ru	Sd to Rd	Sd to Un	Ru to Rd	Ru to Un	Rd to Un
% of stride	24	22	20	20	18	100	31	84	87	84	23	99	62	53	74
Feature	Free resultant	NPP Free resultant	NPP resultant	NPP resultant	NPP X	NPP Free resultant	Free X	NPP Free Z	NPP Free X	NPP Free X	Free Z	NPP Free resultant	Free resultant	Resultant	Z
SMD	1.63	1.66	0.78	0.82	0.66	0.80	1.40	1.12	0.95	1.57	1.06	0.89	0.46	0.46	0.61

Table E.4 Results at highest SMD mid-shank. F=Flat ground, Su=Stair ascent, Sd=Stair descent, Ru= Ramp ascent, Rd=Ramp descent, Un=Uneven terrain

Terrain	F to Su	F to Sd	F to Ru	F to Rd	F to Un	Su to Sd	Su to Ru	Su to Rd	Su to Un	Sd to Ru	Sd to Rd	Sd to Un	Ru to Rd	Ru to Un	Rd to Un
SMD	1.63	1.76	0.78	0.82	0.66	0.80	1.57	1.25	1.03	1.76	1.06	1.02	0.53	0.55	0.62
% of stride	24	86	20	20	18	100	22	34	85	22	22	83	63	77	77
Feature	Free resultant	NPP Free resultant	NPP resultant	NPP resultant	NPP X	NPP Free resultant	NPP Free Z	NPP X	NPP Free Z	NPP Free Z	Free Z	NPP Free X	X	Z	Z

Table E.5 Results at lowest p values knee. F=Flat ground, Su=Stair ascent, Sd=Stair descent, Ru= Ramp ascent, Rd=Ramp descent, Un=Uneven terrain

Terrain	F to Su	F to Sd	F to Ru	F to Rd	F to Un	Su to Sd	Su to Ru	Su to Rd	Su to Un	Sd to Ru	Sd to Rd	Sd to Un	Ru to Rd	Ru to Un	Rd to Un
% of stride	99	84	38	13	88	100	99	99	99	84	3	79	80	53	51
Feature	NPP Free Z	NPP Free resultant	NPP Free Z	X	NPP resultant	NPP Free Y	NPP Free Z	NPP Free Z	Free Z	NPP resultant	Z	NPP resultant	X	Free Z	NPP Y
SMD	1.53	1.47	0.50	0.61	0.47	0.74	1.84	1.28	1.21	1.35	0.92	0.86	0.57	0.49	0.44

Table E.6 Results at highest SMD knee. F=Flat ground, Su=Stair ascent, Sd=Stair descent, Ru= Ramp ascent, Rd=Ramp descent, Un=Uneven terrain

Terrain	F to Su	F to Sd	F to Ru	F to Rd	F to Un	Su to Sd	Su to Ru	Su to Rd	Su to Un	Sd to Ru	Sd to Rd	Sd to Un	Ru to Rd	Ru to Un	Rd to Un
SMD	1.53	1.53	0.50	0.63	0.50	0.74	1.96	1.28	1.21	1.42	0.92	0.86	0.57	0.50	0.50
% of stride	98	85	38	10	89	100	98	98	99	86	86	85	62	29	47
Feature	NPP Free Z	NPP Free resultant	NPP Free Z	X	NPP Free Z	NPP Free Y	NPP Free Z	NPP Free Z	NPP Free Z	NPP Free resultant	NPP Free resultant	NPP Free resultant	X	X	X

Appendix F

Recruitment poster

F.1 Recruitment poster

Lower Limb Prosthetic Patients Required

We are running a study looking at developing a sensor system to assess a lower limb prosthetic users activity levels. The system aims to support prosthetists in their prescription of prosthesis components to better meet patients' needs.

The study will small, non-intrusive involve sensors being attached to your body and prosthesis and measurements being taken as you traverse different terrain. Video or motion tracking data will also be gathered to validate the sensor data.

We need lower limb prosthetic patients that are comfortable with climbing/ descending stairs/ramps, walking on uneven ground, have experience using a walking aid and do not regularly participation in an active sport.

If you would like to take part in the study please contact the research team.

Email: m.wassall@edu.salford.ac.uk

Tel: +44-(0)161-2952679, the phone will likely go to voice mail where you can leave your details that the researcher can contact you on.

Appendix G

Modified QualSyst Tool

G.1 Modified QualSyst Tool

Criterion	"YES" = 2	"Partial" = 1	"No" = 0
C1: Question Objective	The question and the objective of the study are clearly mentioned.	The question and the objective of the study seems not clear	The question and the objective of the study are not provided.
C2: Study design	The study design is appropriated to the question/objective		The study design is not appropriated to the question/objective.
C3: Subjects characteristics	The following parameters are given: Healthy volunteers: number of volunteers, gen-der, mean and SD for the age, height, and weight. Otherwise: number of volunteers, gender, inclusion/exclusion criteria, mean and SD for the age, height, and weight.	The following parameters are given: Healthy volunteers: number of volunteers, gender, mean without SD for the age, height, and weight. Otherwise (2 options): o number of volunteers, gender, inclusion/exclusion criteria and mean without SD for the age, height, weight, number of volunteers, gen-der, mean and SD for the age, height, and weight. Inclusion/exclusion criteria are not given.	Data are missing compared to "Partial".
C4: Experimental protocol	The following parameters are given: Studied locomotion tasks; Walking speed; Transitioning leg (if applicable);Number of trials / locomotion task	The following parameters are given: Studied locomotion tasks. One of the following parameters are given: Walking speed; Transitioning leg (if applicable); Number of trials / locomotion task	More parameters are missing compared to partial.
C5: Critical Tim-ing	The critical timings for each transition are given (if applicable).	The critical timings are given but precisely for each transition (e.g. critical timing occurred at foot contact on the new locomotion mode or at foot off of the previous locomotion mode) (if applicable)	The critical timings are not provided, even though the transitions are studied.

C6: Filter	The filters implemented for each signal are given with the corresponding parameters (e.g. Low-pass 4th order Butterworth filter with a 10 Hz cutoff frequency).	The filters implemented for at least one signal are given with the corresponding parameters. Or the filters implemented for all signals are given without the corresponding parameters (e.g. cutoff frequency)	The filters of the signals are not provided.
C7: Analysis windows	For each analysis window, the following information are provided: Beginning and end of each window; Beginning or end of each window and window length. If multiple windows or sliding windows are used, the overlap or the window increment is provided	One information is not provided (window length or window increment or overlap or beginning or end of each analysis window). For instance, the beginning of the window is provided but the end or window length are not provided.	No information concerning analysis window are given.
C8: Features	The feature set is clearly defined. The equations of each feature are provided or given with references.	The feature set is clearly defined but features equations are not given (no references). Or the equations are given but a feature reduction technique is used but the final feature set is not explicitly provided (for instance PCA to reduce the size of the feature set, but the final number of features is not given).	The extracted features are not mentioned. Note that if the raw data of the sensors were fed into the Machine Algorithm, the criterion was rated 2 out of 2.
C9: Algorithms	The tested algorithms are clearly mentioned, the parameters of each algorithm are provided.	The tested algorithms are mentioned.	The tested algorithms are not mentioned.

C10: Evaluation	The evaluation process of each algorithm is provided (e.g. K-fold cross validation with K = 4)	The evaluation process is given but the parameters are not given (e.g. K not provided for K-fold cross validation). As a result, the data split between train/dev/test sets is unclear.	The evaluation
C11: Results	The results for each algorithm are given (mean and standard deviation).	The results for each algorithm are given with-out the standard deviation.	The mean and the standard deviation are not given. Or the mean and the standard deviation are given but the results are not provided for one of the tested algorithms.
C12: Conclusion	The conclusion is supported by the results		The conclusion is not supported by the results. Note that if the results were rated 0 out of 2, the conclusion can still be supported by the results. For instance, the accuracy of the tested algorithms was estimated from graphics readings and the conclusion is sup-ported by those estimations (higher/lower performances).

Appendix H

Artificial cobblestone dimensions

H.1 Artificial cobblestone dimensions

Table H.1 Artificial cobblestone dimensions.

Cobblestone number	Length (mm)	Width (mm)	Height (mm)	Lengthways distance from edge (mm)	Widthways distance from edge (mm)
1	168	61	17	58	15
2	123	85	16	36	63
3	136	82	12	20	42
4	150	66	19	10	16
5	146	84	16	39	16
6	138	61	16	37	23
7	160	83	13	34	12
8	155	90	16	22	22
9	133	75	17	33	59
10	169	63	18	56	31
11	136	89	15	33	59
12	163	86	12	15	14
13	139	87	15	14	10
14	148	77	16	30	12
15	168	79	14	30	19
16	169	74	12	13	25
17	122	71	14	49	58
18	155	64	19	49	27
19	133	73	19	51	63
20	153	74	15	19	35
21	128	76	15	17	18
22	134	80	18	37	40
23	150	85	19	21	15
24	126	64	19	53	61
25	121	82	18	36	78

26	149	66	15	37	36
27	149	76	12	45	16
28	123	70	18	38	31
29	130	86	13	33	34
30	141	69	19	22	11
31	170	66	14	19	22
32	165	70	14	16	16
33	146	78	15	14	23
34	160	64	16	33	36
35	163	61	18	63	23
36	130	89	13	31	44
37	163	66	19	52	21
38	168	68	18	35	18
39	128	81	15	14	31
40	126	84	19	26	30
41	126	70	13	45	57
42	122	76	17	22	42
43	131	89	12	36	67
44	135	79	17	24	65
45	122	85	17	30	58
46	156	73	12	14	27
47	141	82	14	41	36
48	126	90	16	12	70
49	151	61	18	17	43
50	153	81	16	14	40
51	157	90	16	31	21
52	172	65	17	17	23
53	142	62	18	36	49
54	151	77	12	36	27
55	130	80	18	11	26
56	170	83	15	23	14
57	149	82	13	34	41
58	141	63	17	21	57
59	135	86	14	28	51
60	135	69	19	18	49
61	142	72	15	27	20
62	150	73	19	13	48
63	160	80	13	37	29
64	147	87	13	22	13
65	130	64	12	48	39
66	163	79	16	14	15
67	152	61	17	26	28
68	128	60	13	51	44

69	150	87	14	30	46
70	169	71	15	23	19
71	171	72	17	47	13
72	165	80	19	23	19
73	129	76	19	45	67
74	167	77	13	48	17
75	172	65	19	58	26
76	159	64	13	21	31
77	174	74	14	18	18
78	164	76	13	43	15
79	127	76	18	21	12
80	142	63	18	31	27