Exploring machine learning, real-time bio-feedback, and inertial sensor accuracy for the prevention of running-related injuries

Clare Lillis, B.Sc

A Dissertation submitted in fulfilment of the requirements for the award of

Doctor of Philosophy (PhD)

SCHOOL OF HEALTH AND HUMAN PERFORMANCE

DUBLIN CITY UNIVERSITY

Supervised by Prof. Kieran Moran, Prof. Noel E. O'Connor

August 2022

Declaration

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of Doctor of Philosophy is entirely my own work, that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge breach any law of copyright, and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

Signed:

ID No.: 16213673

Date: 20/08/2022

Acknowledgements

I would like to express my gratitude and appreciation to my two supervisors, Prof. Kieran Moran and Prof. Noel E. O'Connor, for giving me the opportunity to join our research group and for the support and encouragement throughout my PhD. Kieran, I will always appreciate the guidance and patience you showed me over the years. Your passion for research and the work that you do to encourage those around you is inspiring. Noel, I will always be grateful for your support and advice along the way.

I would like to include a special mention to Dr. Amin Ahmadi and Dr. Graham Healy for their technical advice during the PhD. Amin helped me to get the research going, and Graham helped to close it out.

I would like to express every thanks to my friends and colleagues at DCU, for their support and friendship. Special mention to everyone in the labs in L1.29 and L1.30 over the years and the biomechanics trio, who really made our DCU community special for me. To my friend Ghanashyama, finishing the PhD at home was tough without having you to chat every day. I'll always remember our craic at DCU and your friendship over the years.

I will be forever grateful to my parents, Eilish and John, who have supported me like crazy throughout this journey, for always believing in me. It is thanks to every ounce of your encouragement and support that I have this achievement today. To the rest of the family, Caroline, Lawrence and Louise, thank you for being there for me through it all and for making me laugh when things felt tough. This was truly a team effort.

Thank you to my friends, for checking in and provided the space for me to forget about my PhD when I needed it.

Finally, to my boyfriend Damon, your love and support is the reason I was able to make it to the end. Thank you for always being there for me, and always being ready to celebrate a little win as well as a big one.

Contents

List of Poster Presentations

1. Clare Lillis. 'Determining Accuracy of Inertial Sensor to Measure Running Technique for Long Durations'. *The 5th Insight Student Conference* (INSIGHTSC2018), University College Dublin, Dublin, 2018. (Poster and oral presentation)

2. Clare Lillis. 'Determining Accuracy of Inertial Sensor to Measure Running Technique for Long Durations'. *Insight Demonstrator Event*, National University of Ireland Galway, Galway, 2018. (Poster)

3. Clare Lillis. 'Inertial Sensors and Running Technique'. *Dublin City University Faculty Research Day*, Dublin City University, Dublin, 2019. (Poster and oral presentation)

4. Clare Lillis. 'Soft Run: Android Mobile Application for Bio-Feedback'. *Research Seminar Series*, Dublin City University, Dublin, 2019. (Oral presentation)

5. Clare Lillis. 'An Investigation into the Effect of Data Loss on Sensor Derived Orientation Data'. *The 6th Insight Student Conference*, National University of Ireland Galway, Galway, 2020. (Poster and oral presentation)

List of Abbreviations and Acronyms

AUC   Area under the Receiver Operating Characteristic Curve

ACL   Anterior Cruciate Ligament

CF   Complementary Filter

CSV   Comma Separated Value

EKF   Extended Kalman Filter

IMU   Inertial Measurement Unit

GRFs   Ground Reaction Forces

KF   Kalman Filter

KNN   K-Nearest Neighbour

MIMU   Magneto-Inertial Measurement Unit

MARG   Magnetic, Angular Rate, and Gravity

ROC   Receiver Operating Characteristic Curve

ROM   Range of Motion

RMSE   Root Mean Square Error

SGD   Stochastic gradient descent

SVM   Support Vector Machine

UKF   Unscented Kalman Filter

VALR   Vertical Average Loading Rate

VILR   Vertical Instantaneous Loading Rate

vGRFs   Vertical Ground Reaction Forces

3D   Three Dimensional

6DOF   6-Degrees Of Freedom

9DOF   9-Degrees Of Freedom

List of Figures

List of Tables

Abstract

"Exploring machine learning, real-time biofeedback, and inertial sensor accuracy for the prevention of running related injuries."

Clare Lillis

Recreational running is popular, however, incident rates of running related injuries (RRIs) are very high. Predisposition to injury can be assessed through expensive, laboratory-based biomechanical screening. Wearable wireless inertial sensors offer a potential solution, but accurate orientation data are required. This thesis examined the prevention of RRIs, by aiming to improve sensor accuracy, and investigate applications of biofeedback and machine learning.

This thesis explored improving (magnetometer-free) orientation accuracy during running, through examination of (i) Z-axis de-drifting, (ii) data-loss (iii) and modifications to the Madgwick filter. Despite some accuracy improvements (i, iii), overall errors were unsuitable for running based applications.

Impact loading is associated with RRIs, with thigh angle (quasi-measure of knee-flexion) potentially important in load attenuation. Loading can be altered directly (loading-based biofeedback) or indirectly (technique-based biofeedback), these two types of biofeedback were compared. A mobile phone application was developed providing audio biofeedback to reduce impact accelerations and encourage a 'softer' running technique. Both types of feedback reduced loading at the tibia and sacrum, however, tibia loading reduced better with impact accelerations biofeedback, and sacrum loading with thigh angle biofeedback.

It would be beneficial to identify runners who may be predisposed to injury. Seven supervised machine learning models were developed to identify runners who may be likely to sustain RRIs, using inertial, kinematic and clinical data collected on 150 prospectively tracked runners. These models resulted in weak predictive accuracy (0.58-0.61 AUC). As we cannot identify runners predisposed to injury, all runners must be recommended for injury prevention interventions.

Orientation accuracy was found to be sufficient for relative measures of running technique in the biofeedback app. Future work could investigate biofeedback app use in relation to reduction of RRIs. Additionally, running injury prediction could be examined further with respect to extracting different features (continuous measures) or predicting specific injuries.

Chapter 1

Introduction

Running is extremely popular and has proven health benefits [4]; however, running injuries are extremely common with up to 80% of all recreational runners becoming injured each year [5]. The challenges to reducing injury rates in runners is to identify the underlying causes of injury, and to provide a mechanism by which people can subsequently change their running technique to reduce these causative factors. In order to be able to do this we need to develop technologies which allow us to accurately determine an athlete's running technique, modify running technique, and assess the risk of running related injuries (RRIs) due to a person's loading and/or running technique. This research thesis is divided into three areas:

(i) To determine and improve the accuracy of 6-degrees of freedom
orientation estimation using inertial sensors during running.

(ii) To develop a mobile solution to provide real-time biofeedback to reduce
impact loading during running.

(iii) To develop a machine learning model for the prediction of running
related injuries using running impact loading and running technique data.

1.0.1 Determining and Improving the Accuracy of 6-Degrees of Freedom Orientation Estimation using Inertial Sensor during Running

Measuring human movement is very important with regard to identifying inappropriate movement techniques which may predispose to injury e.g., poor running technique. Motion analysis systems are traditionally used to analyse human movement technique with very high levels of accuracy. There are two main drawbacks associated with these systems: their cost and location constraints. These systems are generally limited to a laboratory environment, reducing the available capture space and requiring expensive specialised equipment. These factors in combination with the lack of access

suggests that few people can avail of the potential benefits provided by the highly accurate systems. In addition, capturing movements in a laboratory environment does not allow for the analysis of a runner's true movement patterns which would be demonstrated in their natural running environment. A possible solution is the use of wireless inertial measurement units (IMUs) as an alternative approach to capture human movement data. Inertial sensors provide a portable, low-cost technology, which allows data capture and analysis in a non-laboratory setting. Wireless wearable inertial sensors containing accelerometers and gyroscopes (measures of loading and quasi-technique) have a real potential to become the standard way to collect accurate human movement data.

The work in this thesis focused on a 6-degrees of freedom (6DOF) orientation estimation approach, where 6DOF refers to the use of accelerometer and gyroscope sensors only. A popular approach for orientation estimation is to use 9DOF solutions which also include the magnetometer signals; however, the magnetometer sensor can become disrupted due to unpredictable magnetic distortion and changes in the local magnetic field [6,7]. The magnetic distortion is due to hard and soft iron effects which can cause errors in the calculated orientation. Although compensation methods and calibration procedures are available to limit the effects of magnetic distortion, it is not always possible for the magnetometer to be calibrated in the area of its intended use, especially if the runner is passing through areas with differing local magnetic fields, especially indoors [6,8]. Additionally, most of compensation techniques typically only allow for temporary disruption over short periods of magnetic distortion [9].

It is clear that there are issues associated with the use of 6DOF orientation estimation for running, which relies on the accelerometer signal alone to offset errors in the gyroscope orientation. 6DOF orientation is better able to estimate the gravity vector during slow movements (and therefore orientation estimates) when acceleration is mainly due to gravity [10, 11]. However, the possible magnitude of inaccuracies arising from 6DOF orientation estimation during running is unclear from the literature. This makes it difficult to conclude whether 6DOF orientation estimation could still be useful and sufficiently accurate for some applications of measuring running technique (i.e., where only comparing changes in relative measures [of a particular subject] of orientation/joint angles are needed).

There are three main limitations to calculating accurate 6DOF running orientation or joint angle data from inertial sensors data: signal drift [12], movement type (impacts and speed [8]) and data loss [13]. The accuracy of calculating orientation, joint angles and position estimates derived from inertial sensor data are affected by signal drift [12]. Signal drift is a phenomenon which occurs when

the noise and bias associated with sensor data (e.g., accelerometer and gyroscope) causes estimates of orientation (calculated by gyroscope data integration) to become more erroneous over time. On examination of related studies, the Madgwick filter [14] is one of the most commonly applied approaches to date to account for this issue of signal drift. This is the baseline orientation algorithm used for the investigation of segment orientation inaccuracy during running in Chapter 7. For inertial sensors to be used in 'free-living' conditions, the sensors could be required to capture data accurately from a couple of minutes to a number of hours. This is especially true if the sensors are being worn to capture movements in a training environment, where accurate data over a long period of time would provide more interesting insights. For this reason, movement data were collected over a duration of >30 minutes to assess the orientation accuracy over time.

Research would appear to have mainly focused on slow and dynamic movements for testing 6DOF orientation algorithms, with less research quantifying the effect of high speed, impact related movements, such as running. Studies have shown the clear relationship between faster movement speeds and higher orientation error [8, 15, 16]. This increase in error is likely caused by body acceleration (accelerations other than gravity, when an accelerometer is attached to a moving object/person) creating disturbances in orientation estimates [10, 11]. However, it is still important to be able to quantify the level of error in different movements and over time, especially with regard to the non-drifting axes (pitch and roll), which may not be affected by signal drift but will be affected by the external accelerations. Some studies focus on using inertial sensors to examine other aspects of running (e.g., estimation of running speed [17] or running gait symmetry [18]). For this reason, our studies have involved the collection of both running and walking data to directly examine the effect of impact and movement speed on the algorithm accuracy.

Data loss can affect all sensors and can occur during normal sensor usage or due to technical issues. As estimating current orientation depends on knowing the previous orientation, there is an intuition that the previous orientation should be accurately known in order to confidently estimate the current orientation. It is possible that inaccuracies in orientation due to data loss may have a knock-on effect on the accuracy of successive orientation estimates [13]. The first study in-part examines the effect of data loss on the 6DOF orientation estimation of body segments during running.

It may be possible to boost the running orientation accuracy by experimenting with changes in the orientation estimation algorithm. A number of studies have examined the tuning of the beta parameter within the Madgwick filter in order to provide better orientation estimates [12, 19]. This

will be explored for our 6DOF running orientation to distinguish whether this could be helpful for our application to obtain sufficiently accurate orientation. In addition, a different algorithm modification was investigated where the main Stochastic Gradient Descent block within the algorithm was replaced with a Stochastic Gradient Descent Optimiser, momentum. This has not been examined in any studies to-date.

1.0.2 Developing a Mobile Solution to Provide Real-time Biofeedback to Reduce Impact Loading during Running

Given that loading on the body and the associated technique are thought to be related to the occurrence of running related injuries [20], it would be beneficial to be able to perform running technique re-training. Running technique re-training would facilitate runners in reducing loading on the body while they run [21]. The challenge with this is that running is a highly repetitive action that appears to be generally controlled subconsciously [22]. In order for runners to adjust their running technique, it has been suggested that near real-time biofeedback is required [23]. This feedback can be technique-based or loading-based, but it is unclear which form of feedback may be more beneficial. Only one study by Baggaley et al. appears to have directly compared these approaches to reduce vertical ground reaction forces during running [24], and to-date it would appear that no studies have compared the two types of biofeedback for the reduction of impact accelerations measured via accelerometers.

In addition, as described above, the issue of the accuracy of 6DOF measures for providing technique-based biofeedback in real-time in a natural running environment needs to be examined, as no studies appear to have provided knee kinematics based biofeedback as estimated via inertial sensors during running. The provision of biofeedback in a natural running environment is especially important as although there are a large number of studies which focus on running re-training, the application of findings from previous research is somewhat limited due to the majority of studies focusing on the provision of biofeedback through the use of expensive equipment in constrained laboratory settings with limited access [25]. Therefore, it is important to assess the feasibility of providing technique-based biofeedback via inertial sensors for running re-training.

Running re-training literature focuses heavily on the assessment of body loading at the tibia, while largely neglecting assessing other body segments which may also be sites of injury i.e., the sacrum [26]. In this thesis, impact loading (peaks and rates of acceleration) was assessed before and during biofeedback at both the tibia and the sacrum. Therefore, this thesis will examine a comparison of

real-time feedback on 6DOF running technique (thigh angle) versus real-time feedback on impact loading (impact accelerations) in an outdoor running environment. The aim was to examine whether impact acceleration-based biofeedback (loading) or thigh angle-based biofeedback (technique) provided greater reductions in loading measured at the tibia and sacrum.

1.0.3 Developing Machine Learning Models for the Prediction of Running-Related Injuries using Impact Loading and Running Technique

There are conflicting findings with regards to the bio-mechanical causes of running related injuries [27–30]. The literature suggests that the magnitude and rate of loading on the body may be important [31, 32] as well as elements of running technique which affect these magnitudes of loading [29, 33]. One of the challenges facing the literature to date is that it has been retrospective in general, and where loading on the body has been assessed it has been at a whole-body level using ground reaction forces. Injuries do not occur at at a whole body level, therefore segmental accelerations may be more important in predicting running injuries. Large-scale prospective studies are needed in which both factors related to loading on the body and running technique are taken into account.

Many studies look to identify relationships between variables (e.g., intrinsic or extrinsic risk factors) and injury [34, 35]. This can be important as some identified relationships can be targeted for modification in an injury prevention intervention, in order to reduce the effect of the identified risk factor. However, it can be difficult to obtain full participation from runners in running retraining strategies [36], and so it would be extremely beneficial to be able to identify those runners predisposed to injury to target for injury prevention strategies. If a runner knows that they may be at risk of incurring an injury, it could significantly motivate them to participate in an injury prevention programme. In this thesis, we aimed to build machine learning models to identify those predisposed to running related injuries. Data were analysed from 150 runners who were prospectively tracked for running injuries for 1-year post data collection. These data were used to create datasets with prospectively injured and uninjured runners. The models were trained using a mixture of sub-datasets comprised of inertial data (accelerometer and gyroscope), kinematic technique (joint angles) and clinical variables. This thesis examined in the first instance, whether the machine learning predictions were sufficiently accurate, and additionally whether there was any added improvement in prediction by including the supplementary information on running technique via joint angles as opposed to inertial variables alone (measure of loading and quasi-technique).

Seven classifiers were trained on each sub-dataset and their performance compared.

1.1 Research Objectives and Hypotheses

Research Objectives

Study 1

- **Aim 1.1:** *To determine the magnitude of inaccuracy in calculating segment orientation using baseline 6DOF algorithm during running with novel drift compensation, by comparing orientation estimated via inertial sensors (accelerometer and gyroscope) to a reference optical motion analysis system.*

- **Aim 1.2:** *to evaluate the performance of creating a gap filling pre-processing pipeline of pre-filling and post-filling, for improved orientation accuracy.*

- **Aim 1.3:** *To evaluate the performance of using modified versions of the baseline algorithm, in order to determine the effectiveness of the modified approaches compared to the baseline accuracy.*

Study 2

- **Aim 2.1:** *To develop a mobile phone application with the capability to provide real-time biofeedback based on impact accelerations (impact loading) or segment orientation (thigh angle technique).*

- **Aim 2.2:** *An investigation of which of these two types of biofeedback are most effective in reducing impact loading at the tibia and sacrum during running.*

Study 3

- **Aim 3.1:** *To compare different datasets using inertial data (accelerometer and gyroscope), movement technique data (joint angles) and clinical data in determining which combination or dataset is more effective at identifying those who are predisposed to injury.*

- **Sub-aim 3.2:** *To identify the injury causative factors related to running when using measurements of technique and impact acceleration.*

- **Aim 3.3:** *To determine which machine learning classifier is more effective at identifying those who are predisposed to injury.*

Research Hypotheses

From the research questions listed above, the author outlines the following hypotheses.

Study 1

H1. 6DOF inertial sensor orientation will be less accurate than the Vicon system data for calculating segment orientation.

H2. Filling missing samples in the inertial data (before calculating orientation) will lead to more accurate orientation data.

H3. The modified implementations of the baseline algorithm will provide an improvement in orientation accuracy over the baseline implementation.

Study 2

H1. Both loading-based (impact accelerations) and technique-based (thigh angle) biofeedback will result in a reduction in loading on the tibia and sacrum during running, but loading-based biofeedback will be more effective.

Study 3

H1. Using a combination of impact accelerations and movement technique will be most effective at identifying those who are predisposed to injury.

H2. Certain factors will be identified as predictive factors of running injury using machine learning.

1.2 Thesis Outline

This thesis contains three main studies (as described above) which are linked through the broad frame of using loading and running technique-based measures with to design mechanisms for the prevention of running related injuries. It is important to highlight to the reader that although this link exists across studies, each study is stand-alone in the examination of specific areas. This approach was chosen as it was important in the first instance to overcome challenges to improve the level of 6DOF orientation accuracy (Study 1) in order to design studies which would rely on accurate segmental loading and technique data. Sufficiently accurate segment/joint angle data would allow the assessment of running technique, and the provision of accurate biofeedback to alter running technique (Study 2) or to build machine learning models capable of distinguishing those predisposed to injury (Study 3). Both approaches which potentially have the ability to reduce the risk of RRIs. The chapters in this thesis are organised as follows.

Chapters 2 - 5 provide an extensive overview of the background information and related work most relevant to the different problem areas of this thesis. These chapters include: an introduction to running kinematics and kinetics, the risk factors of running related injuries, an overview of orientation calculation methods, the use of biofeedback in running, and an introduction to machine learning.

Chapter 6 describes each of the three main datasets used for the three distinct areas investigated in this thesis. This chapter provides an overview of each experiment, along with information on the equipment used, participants recruited, and the data collection protocol followed.

Chapter 7 outlines the first study and set of experiments, examining the magnitude of inaccuracy associated with 6-degrees of freedom orientation estimation, and a number of following experiments aiming to improve the level of orientation accuracy to an acceptable level.

Chapter 8 outlines the second study, in which a mobile phone application was developed in order to investigate the provision of biofeedback to runners in an outdoor environment in real-time. Two types of biofeedback were examined with two groups of runners, where biofeedback provided to one group was based on loading at the tibia (impact accelerations) and biofeedback provided to the other group was based on lower body technique (thigh angle at initial contact).

Chapter 9 outlines the third and final study, in which a master dataset was collected comprised of inertial, kinematic and clinical data collected on prospectively injured and uninjured runners. The aim of this study was to use the prospective data to build a machine learning model capable of identifying

those runners predisposed to injury. The main dataset was split into six sub-datasets (inertial data only, kinematic data only, clinical data only and all combinations of inertial, kinematic and clinical data). Each of these 6 sub-datasets were used to train 7 different machine learning classifier models, in order to determine whether kinematic [joint angles (technique)] data were beneficial to the prediction task compared to inertial data alone [accelerometer and gyroscope (measure of loading and technique)], and which machine learning classifiers performed the best prediction. The aim was to be able to use the best model to identify injury causative factors for running related injuries.

Chapter 10 provides a summary of the work carried out in this thesis, along with a discussion of the results with respect to the aims of each study.

Chapter 2

Related Work: Kinematics, Kinetics and Running Injury

2.1 Running Kinematics and Kinetics

This section serves as an introduction to ground reaction force and the running gait cycle. A brief description of the kinematics and kinetic of running follows which will allow future sections to discuss how these factors may be related to injury.

2.1.1 Ground Reaction Force

When the foot strikes the ground during running the ground exerts an equal and opposite force back on the body [37]. The force exerted by the ground is known as the ground reaction force (GRF). These high forces travel as a shock wave through the body and have been associated with running related injuries [20], as injuries are caused by high loading relative to the body's tissue strength [38]. The vertical ground reaction force (vGRF) curve, particularly for a rear-foot runner, has two distinct peaks [termed the impact peak (passive peak) and the propulsive peak (active peak)] and a rate of loading [1]. Running related research tends to focus on the effect of the impact peak, and the loading rate of vGRF (see figure 2.1) during running, with less emphasis on the role of the active peak. This is because of the possible relationship between the impact phase and injury [20], and due to the absence of an impact phase in some runners [39], the loading rate may provide valuable information with respect to the loading after initial contact [40]. Vertical GRFs represent a measure of loading on a whole-body level, as the total force is calculated as a sum of its segment parts. Therefore, vGRF is not segmentally proportional, there is no information about how the forces are distributed across body segments. For example, two individuals with the same measure of loading on the whole body may have different magnitudes of loading at individual body segments.
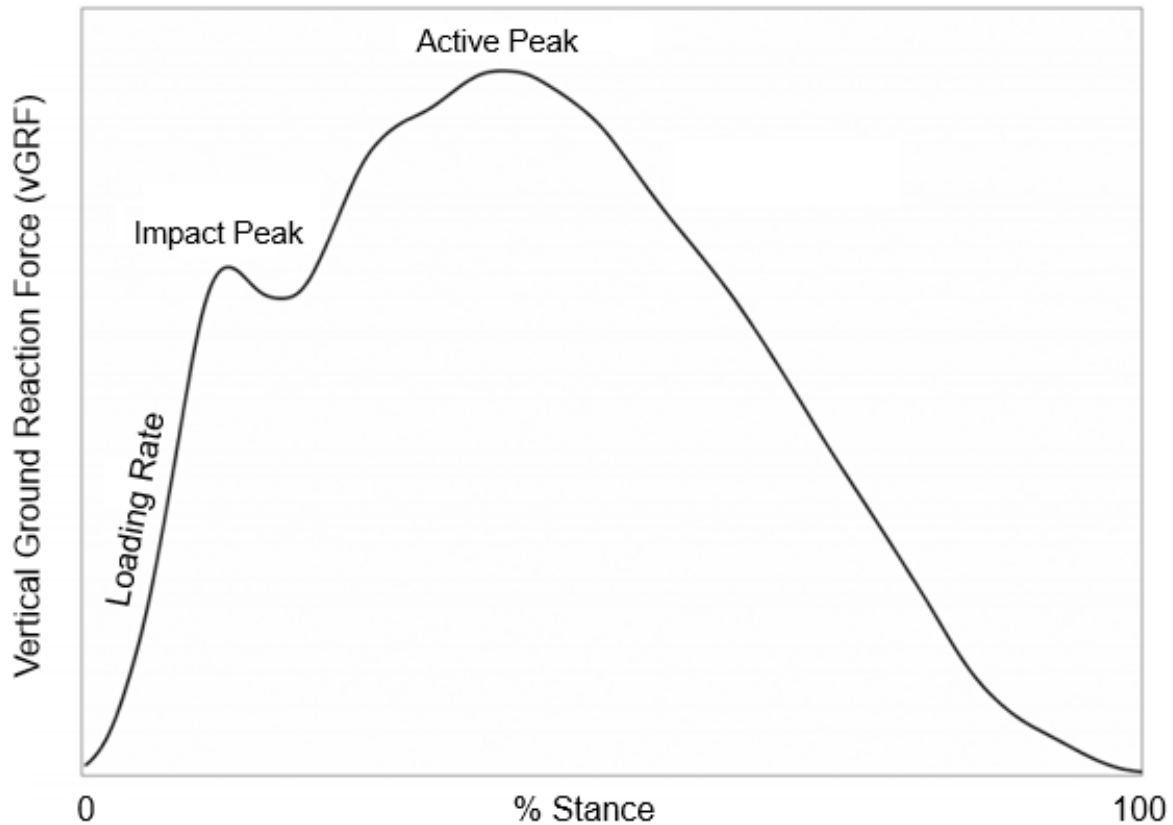
Figure 2.1: The curve depicts vertical ground reaction force as a percentage of stance phase during rear-foot running, showing the two main distinctive peaks: the impact peak and the active peak. Loading rate is calculated as the gradient of the ground reaction forces occurring before the impact peak, representing the speed at which force is applied to the body. Figure adapted from Van et al. [1]

The impact peak appears approximately within in the first 15% of the stance period [40], lasting up to approximately 10-30ms after initial contact with the ground [41]. The approximate magnitude of these impact forces may range from 1.5 to 5 times the person's body weight depending on the speed of movement and landing geometry [41]. The impact peak is followed by the active peak which occurs at mid-stance approximately between 35- 50% of the total stance time [40], with active forces lasting up to 200ms [41]. The loading rate describes the speed at which (impact) forces are applied to the human body. Loading rate is represented by the linear gradient of the vGRFs curve from initial contact to the impact peak, and commonly reported using vertical instantaneous loading rate (VILR) [42] and vertical average loading rate (VALR) [42]. VALR has been defined as the slope of the line from the start to the end of the time-period where the GRF curve is linear. VILR is the maximum slope of the vGRF curve between consecutive points at the same time-period [43]. There are a number of techniques used to calculate these loading rates, but the majority of studies calculate loading rates

over the range of stance time where the vGRF is linear [44]. Different techniques to calculate loading rates differ in how they define the time period over which loading rates are calculated. This time-period can be defined in different ways: as the (stance) time from foot strike to first peak, from 20 to 80% of the stance time to first peak, or over the first 50ms after initial contact [42].

2.1.2 Acceleration

A large number of studies have investigated ground reaction force with respect to injury (see table 2.3). The challenge however is that GRFs capture a summed measure of loading on the whole body as opposed to a measure of loading at individual segments. Furthermore, measuring vGRF is not segmentally proportional, there is no information about how the forces are distributed across body segments.

$$GRF = \Sigma(mi * ai) \tag{2.1}$$

where mi is mass of each segment, and ai is acceleration of each segment.

Given that injuries occur locally at individual segment level, it may be more beneficial with respect to injury to assess loading locally at individual body segments. One approach that has been commonly used to measure segmental loading is the use of wearable accelerometers that can be attached to individual body segments [21, 27]. Accelerometers can be used to measure instantaneous peaks and rates of acceleration in real-time. Accelerometers not only allow for a segmental measure of loading, but also allow for loading-based data capture to occur outside of a laboratory environment. Accelerometers have most commonly been used to measure tibial impact accelerations, as the lower limbs are commonly affected by running related injuries [20].

2.1.3 The Running Gait Cycle

Running has increased velocity in comparison to walking and incorporates an aerial phase (double float phase), which is a point in time when no limbs are in contact with the ground [45]. The gait cycle is a repeated pattern, starting when one foot makes contact with the ground (initial contact) to the next initial contact of the same foot [46].

The running gait cycle can be divided into stance phase, swing phase and float phase. The cycle begins with the stance phase, the first half of which involves absorption (initial contact to mid-stance)

and the second half involves propulsion (mid-stance to toe-off). The stance phase is followed by the swing phase which is subdivided into the initial swing (toe-off to mid-swing) and the terminal swing (mid-swing to initial contact) [2]. The double float phase occurs at the start of initial swing (after toe-off) and again at the end of terminal swing (before initial contact)[47]. The double float phase is the moment at which neither foot is touching the ground.
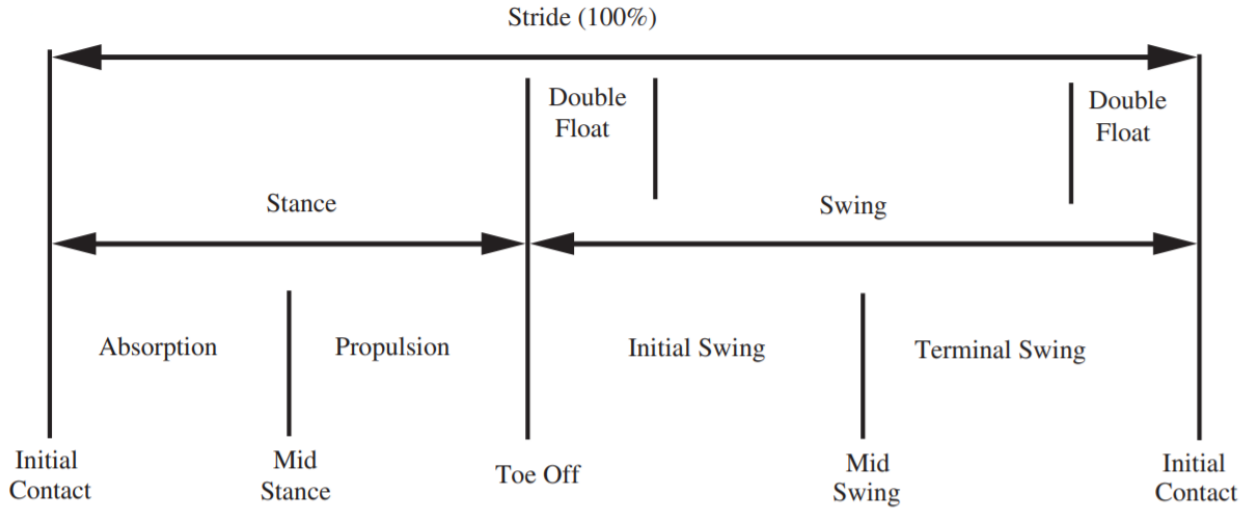
Figure 2.2: The running gait cycle for one stride, showing the multiple phases which occur during one running gait cycle. Figure adapted from Dugan and Bhat [2]

2.1.4 Running Kinematics

Running gait kinematics are detailed in the following section with respect to the primary joints and their primary actions.

2.1.4.1 Foot and Ankle Kinematics

Supination and pronation are a combination of sagittal, frontal and transverse plane movements of the talocrural and subtalar joints of the ankle involved in running gait. Movements concerning the sagittal plane through the talocrural joint specifically are plantar flexion and dorsiflexion [2]. With reference to a rear-foot running pattern, the foot moves to a greater dorsiflexed position at initial contact as the heel makes contact with the ground [46], which has been reported in the range of 3-6 degrees dorsiflexion [48, 49]. The foot moves to a pronated pose within the first 20% of the stance phase, in order to absorb some of the impact shock at heel strike [2, 46]. The shock absorption lasts until the foot reaches its maximally pronated position. The ankle has an average range of motion (ROM) of 45 degrees [2]. Towards the end of the stance phase, the runner pushes off the ground as the foot

moves into plantar flexion reaching approximately 25 to 35 degrees at toe-off [2]. As the foot moves through the swing phase, the foot position changes from plantar flexion to a dorsiflexed position, before returning to a neutral position, as the running gait cycle pattern restarts with dorsiflexion directly after initial contact [2].

2.1.4.2 Knee Kinematics

Increased knee flexion has an important role in increasing shock attenuation at foot strike [46]. The level of knee flexion during initial contact ranges from approximately 15 degrees to 25 degrees [48–50]. As motion continues into the mid-stance phase, knee flexion ranges from 38 degrees to 45 degrees [46, 48, 51]. As the knee moves through the propulsion stage, the knee extends to approximately 25 degrees [46], as the action of knee extension aids the body to push forward to the running double float phase (see figure 2.2). The knee flexes again into the swing phase before finally extending again into the next initial contact foot-strike [46, 48].

2.1.4.3 Hip and Trunk Kinematics

Hip flexion angles can range from approximately 20-50 degrees of flexion at initial contact [46–48]. After foot-strike at initial contact, the hip extends and can reach magnitudes of hip extension of approximately 4-11 degrees [47] once maximally extended. The hip stays in an extended position until the mid-swing phase when the contralateral foot strikes the ground and the hip reverses direction into flexion (ranging from approximately 30-65 degrees) [47, 48]. Once maximum flexion is reached approximately around the time of contralateral toe-off, the hip extends again in anticipation of initial contact [47]. The trunk comprises of approximately 50% of the body's mass, and therefore has a significant influence on the loading of the body as a whole. The trunk is up-right at initial contact and flexes during the stance phase with mean trunk flexion of approximately 4-11 degrees [47, 51]. During initial contact, the lower limbs in particular must act to absorb loading associated with deceleration of the trunk.

2.1.5 Running Kinetics

2.1.5.1 Foot and Ankle Kinetics

The main role of muscles are to perform shock absorption during the occurrence of impacts at initial contact, to assist posture control of the body, to generate energy for forward propulsion and to manage

direction changes of the body's center of mass [52]. With reference to a rear-foot running pattern, before the heel strikes the ground at initial contact the anterior tibial muscles eccentrically contract, lengthening the muscle as the forefoot reaches the floor [46]. During pronation of the foot, the muscles of the lower leg and foot eccentrically contract (lengthening the muscle) in order to stabilise the ankle joint after impact and provide shock absorption [2]. The ankle joint musculature converts the energy absorbed during mid stance into generation of power to supply energy to propel the body forward [46]. This occurs during the supination phase where the muscles concentrically contract in order to shorten the muscle fibres in preparation of this propulsion phase and associated acceleration stage [2]. The foot moves into the swing phase as anterior tibial muscles work to create ankle dorsi-flexion to avoid non-heel ground contact. This is followed again by eccentric contraction of muscles to control lowering of the fore-foot, as the ankle returns to a more neutral flexed position again [46], before increased dorsiflexion directly after initial contact, restarting the gait cycle [2].

2.1.5.2 Knee Kinetics

The amount of knee extension which occurs just before initial contact is constrained by the eccentric contraction of the hip extensor muscles specifically the hamstrings [46]. The knee moves into a flexed position at initial contact in order to reduce the effect of the ground reaction forces on the body. The impact of the ground reaction force is absorbed by the eccentric contraction of the quadriceps as the knee flexes [46]. As the knee extends in the stance phase, the quadriceps contract concentrically generating power which is used as energy to propel the body forward. From this position the quadriceps control knee flexion in the swing phase until the late swing phase where the knee extends to prepare for initial contact. In the swing phase, as the knee maximally extends, the knee muscles absorb power rather than generating it, in order to control the leg during the swinging movement. To avoid hyper-extension of the knee as the leg is preparing for initial contact, the hamstring muscles contract eccentrically to control the movement [46].

2.1.5.3 Hip Kinetics

The amount of hip flexion which occurs just before and at initial contact is constrained by the eccentric contraction of the hip extensor muscles (gluteus maximus, hamstrings). The hip flexor and extensor muscles are responsible for power generation. These muscles then concentrically contract as the hip extends in the stance phase generating power. Maximum hip extension is quickly followed by hip

flexion as the leg moves from the later stance phase (toe-off) through to the swing phase. The hip extensors contract concentrically to allow the hip to move from peak flexion into extension at the end of the swing phase as the hip returns to its pre-initial contact position [46].

2.2 Running-Related Injury

2.2.1 The Relationship between Loading and Injury

Injuries occur at a tissue level, due to the accumulation of tissue damage and progressive loss of stiffness and strength due to repetitive loading and frequent activity [53]. Repetitive loading causes bouts of micro damage, which can positively adapt given the proper rest and repair [41,53]. Without allowing for repair, continued loading leads to accumulation of damage which could cause injury [41,53]. This type of relationship can be described by a stress-frequency curve, which shows the relationship between the stress magnitude applied to a structure and the frequency of stress instances applied as seen in Hreljac [54]. Injuries occur due to excess loading relative to tissue strength [41,53]. In an animal study performed on bovine cartilage explants [55], Ewers et al. found higher rates of loading were associated with greater cartilage matrix damage, with a greater total length and average depth of surface lesions, compared to explants subjected to a low rate of loading. In addition significantly greater amounts of stress were generated at strains experiencing the higher rate of loading compared to the low rate of loading [55]. A model of running injury by Berteslen et al. [56] describes the aetiology of running injuries. The factors highlighted are: load capacity, loading frequency and the amount of load applied (magnitude and distribution of loading).

With respect to load capacity and frequency, structure-specific load capacity is defined as *the musculoskeletal system's ability to withstand load without sustaining injury* [54]. Each body structure of a runner has a unique structure load capacity associated with it. Structure load capacity can change due to a number of various factors, for example, training regularity [57], sex [58], age [59]. When load exceeds the capacity of a given structure [54] injuries can occur. This instance of exceeded load capacity might happen at one event e.g., race or training session, or gradually accumulate strain over multiple running events e.g., series of training sessions, reducing the structure specific load capacity if the appropriate recovery is not complete [56,60]. It is the act of the frequency of loading which can cause the progression of diminishing load capacity, which can continue to the point where risk of injury increases as the load capacity of a specific structure is exceeded [54]. Frequency in this sense

can have a number of interpretations: the number of repetitions of loading (e.g., cadence), the time between each repetition (e.g., stride length) and time between loading sessions (e.g., training sessions) [54].

With regard to the amount of loading applied, load distribution per stride is described as *how the load per stride is distributed across individual anatomical structures* [56]. The distribution of loading can be influenced by changes in technique or simply changing running footwear. In addition to load distribution, magnitude of loading is a factor that has gained a lot of attention as it is suggested that greater magnitudes of loading are related to injury. A study by Edwards et al. [53] showed increased loading cycles resulted in moderate increases in risk of injury for tasks with low magnitudes of loading, but rapid increases in risk for tasks with high magnitudes of loading. This is supported by a number of retrospective studies [27, 32, 61, 62], where the previously injured subjects were shown to have greater loading variables compared to uninjured subjects. Reducing the magnitude of vertical impact peak loading in via technique has been shown in Chan et al. [20] to reduce the magnitude of both impact and rate loading variables (VIP, VALR, VILR) and additionally reduce the risk of injury.

2.2.2 Definition of Injury

Running is a very popular form of exercise due to the large number of health benefits, as well as its lack of location-based restrictions which may be associated with other sports. A disadvantage associated with running is the relatively high rates of running related injuries (RRIs), with incidence rates ranging between 19% and 79% [58]. This large variation can also be seen with reported prevalence, with a recent study on marathon runners reporting the prevalence of RRIs ranged from 29.2% to 43.5% [63]. Differences in the definition of what constitutes an injury across running studies, as well as varying study populations, largely contributes to this wide range in incidence and prevalence rates [64], [65]. Table 2.1 shows a summary of running studies which have included a clear definition of RRIs, and specifically focused on novice and recreational runners. The definition of RRI was split into 3 distinct categories in order to compare across studies. These categories were: region of injury, physical complaint and disruption to training.

Twenty-eight studies were included in the review, with 21 prospective studies (11 studies on recreational runners, 7 studies on novice runners and 3 studies on a mixed population) and 7 retrospective studies (all on recreational runners). With respect to region of injury, 15/28 studies included a specific region of injury in their definition of RRI, with the majority of these studies

reporting either the 'lower extremity', or the 'back or lower extremity' as sites of injury.

All 28 studies included a physical complaint descriptor in their definition of RRI, with 'pain', 'complaint' and 'injury' the most common, appearing in 9/28, 7/28 and 7/28 of the studies, respectively.

With respect to disruption of training, 'lost time' and 'negatively impacted training' were used to describe the training disruption due to injury. Included in the table, 'Lost time' referred to where RRI definitions included a minimum amount of disruption time to training e.g., days or weeks. Similarly, 'Negatively impacted training' includes definition terms which suggested the injury negatively impacted training e.g., lower training intensity or duration due to pain or injury. 26/28 of studies included some training disruption information in their definition of RRIs, with 9/26 studies including only 'Negatively impacted training' in their definition of RRIs, while 17/26 included both 'Negatively impacted training' and 'Lost time'. The definition of lost training time varied across studies from a measure of days, weeks or training sessions lost, which made it difficult to compare across training disruption information of the 26 studies.

Research has shown that stricter definitions are needed to properly investigate RRI incidence rates [64]. Bovens et al. [66] used a broad definition of RRI, "any physical complaint developed in relation with running activities and causing restriction in running distance, speed, duration or frequency" and found an incidence rate of 85%, whereas when a stricter description was used on how long the injury persisted within the definition of RRI, "an injury that cause the runners to stop running for at least seven days", incidence rates dropped to 24% [67]. Following this, a number of research experts established a consistent definition of RRIs to be used in future running injury studies, defining RRIs as 'running-related (training or competition) musculoskeletal pain in the lower limbs that causes a restriction on or stoppage of running (distance, speed, duration or training) for at least seven days or three consecutive scheduled training sessions, or that requires the runner to consult a physician or other health professional' [64].

Table 2.1: Definitions of running related injury across studies

Author	Study Type	Population	Region of Injury	Physical Complaint	Training Disruption
Buist et al., 2010	Prospective	532 Novice Runners	Back or Lower Extremity	Pain	NIT, Lost time(1 week or 3 sessions)
Bredeweg et al., 2013	Prospective	210 Novice Runners	Back or Lower Extremity	Complaint	NIT, Lost time(1 week)
Nielsen et al., 2013	Prospective	927 Novice Runners	Back or Lower Extremity	Complaint	NIT, Lost time(1 week)
Nielsen et al., 2014	Prospective	933 Novice Runners	Back or Lower Extremity	Complaint	NIT, Lost time(1 week)
Kluitenberg et al., 2015	Prospective	1696 Novice Runners	Back or Lower Extremity	Complaint	NIT, Lost time(3 sessions)
Smits et al., 2016	Prospective	185 Novice Runners	-	Pain	NIT, Lost time(1 week or 3 sessions)
Lun et al., 2004	Prospective	87 Recreational Runners	lower extremity	Symptom	NIT
Gerlach et al., 2005	Prospective	87 Recreational Runners	Back or Lower Extremity	Injury	NIT
Van Middelkoop et al., 2008	Prospective	694 Recreational Runners	Lower Extremity	Injury	NIT
Malisoux et al., 2015	Prospective	264 Recreational Runners	-	Pain or Complaint	NIT, Lost time(1 day)
Davis et al., 2016	Prospective	249 Recreational Runners	-	Pain	NIT
Van der Worp et al., 2016	Prospective	417 Recreational Runners	Back or Lower Extremity	Pain	NIT, Lost time(1 day)
Messier et al., 2018	Prospective	300 Recreational Runners	-	Symptom	NIT
Napier et al., 2018	Prospective	65 Recreational Runners	Back or Lower Extremity	Pain	NIT, Lost time(3 sessions)
Dallinga et al., 2019	Prospective	678 Recreational Runners	-	Complaint	NIT, Lost time(1 week)
Taunton et al., 2003	Prospective	844 Novice and Recreational Runners	-	Pain	NIT
Chorley et al., 2002	Prospective	1548 Novice and Recreational Runners	-	Disorder	NIT
Buist et al., 2010	Prospective	629 Novice and Recreational Runners	Back or Lower Extremity	Pain	NIT, Lost time(1 day)
Williams et al., 2001	Retrospective	40 Recreational Runners	Lower Extremity	Injury	NIT, Lost time(1 week)
Tauton et al., 2002	Retrospective	2002 Recreational Runners	-	Pain or Symptom	NIT
Hespanhol et al., 2012	Retrospective	200 Recreational Runners	-	Pain	NIT, Lost time(1 session)
Vadeboncoeur et al., 2012	Retrospective	194 Recreational Runners	Hip or Lower Extremity	Injury	-
Besomi et al., 2013	Retrospective	200 Recreational Runners	-	Distress	NIT, Lost time(1 day)
Ellapen et al., 2018	Retrospective	4380 Recreational Runners	-	Injury	NIT, Lost time(1 week)
Linton et al., 2018	Retrospective	1145 Recreational Runners	-	Problem	NIT

NIT: Negatively Impacted Training.

2.2.3 Biomechanical Risk Factors Associated with Running-Related Injuries

2.2.3.1 Loading Factors that Cause Injury

Injury during running has been suggested to be related to high loading during the impact phase at ground contact [27]. Two measures thought to be important to RRIs are: peak impact and rate of loading [1]. These measures and how they relate to running injury will be explored in the following section.

The Relationship between GRF and Injury

Many studies have investigated the link between loading and injury using the peak impact and rate of loading (VALR and VILR) obtained from vGRF data. Although the area is well researched, findings remain mixed.

With respect to loading rate, 18 studies (6 prospective studies and 11 retrospective studies) have been found to investigate its relationship with RRIs. Of the 6 prospective studies, 3/6 examined both VALR and VILR [31, 68, 69], and 3/6 examined VILR only [65, 70, 71]. With respect to VILR, 2/6 studies found a significantly higher VILR associated with the injured groups compared to the uninjured populations, VILR differences of 31% and 21% were reported by Bredeweg et al. [65] and Davis et al. [31], respectively. Only 1 of the 3 studies focusing on VALR found a significant relationship with RRI [31], with the injured group having significantly higher VALRs compared to the uninjured group (difference in VALR of 29%).

Retrospectively, 3/11 studies examined both VILR and VALR [27, 32, 72], 6/11 studies examined VILR only [28, 62, 73–76] and 2/11 studies examined VALR only [77, 78]. With respect to VILR the findings are mixed, 4 of the 9 studies found significantly higher VILRs associated with the injured groups compared to the uninjured groups [27, 32, 62, 73], with difference in loading rates between injured and uninjured groups ranging from 16% - 46%. In contrast, Duffey et al. [28] found a significantly lower VILR in the injured group when compared to the uninjured group (VILR difference 9%). Four of the five studies focusing on VALR found a significant relationship between VALR and specific RRIs [27, 32, 77, 78], with the injured groups having significantly higher VALRs compared to the uninjured groups (difference in VALR between injured and uninjured groups ranged from 19% - 52%).

The mixed findings reported in the retrospective studies for VILRs could be explained by the difference in subject condition at time of testing. All 4 studies which found a significantly greater

VILR associated with injury had asymptomatic participants at the time of testing, whereas Duffey et al. [28] which found a relationship with significantly lower VILR, tested on participants that were symptomatic which could suggest that they adjusted their running technique in order to minimise loading rates and further reduce knee pain [1]. Messier et al. [75] and McCrory et al. [74] reported similar levels of magnitude in the uninjured group, but the magnitude in the injured group of Duffey et al. [28] was considerably lower in comparison. It is also important to note that the significant finding for both VALR and VILR in Davis et al. [31] was only reported for the sub-group of never injured versus prospective first injury, and no significance was found in the main larger groups of general RRI and uninjured.

Table 2.2: The relationship between loading rate and running related injuries.

Author	Study Type/Loading Rate	Population	Running Injury Time-Frame	Mean(SD) BW/s Uninjured//Injured	Significance (p value, Effect Size)	% Difference
Davis et al., 2016 (sub-group)	Prospective/VALR	Injured: 11, Uninjured: 21 (F)	2 years	60.7 (12.8) // 78.2 (11.1)	P = 0.001*, ES: Large	29%
Dudley et al., 2017	Prospective/VALR	Injured: 12, Uninjured: 19 (X)	14 weeks	58.1 // 68.3	P = 0.313, ES: Small	18%
Napier et al., 2018	Prospective/VALR	Injured : 22, Uninjured: 33 (F)	15 weeks	Not reported	P > 0.05	Not reported
Davis et al., 2016	Prospective/VALR	Injured: 144, Uninjured: 105; (F)	2 years	73.6 (20.7) // 71.3 (18.7)	P =0.357, ES: Small	-3%
Bredeweg et al., 2013	Prospective/VILR	Injured: 34, Uninjured: 176 (X)	9 weeks	96.3 (29.0) // 101.0 (28.4)	P > 0.05	5%
Davis et al., 2016	Prospective/VILR	Injured: 144, Uninjured: 105 (F)	2 years	85.2 (22.7) // 81.1 (20.4)	P = 0.140, ES: Small	-5%
Dudley et al., 2017	Prospective/VILR	Injured: 12, Uninjured: 19 (X)	14 weeks	109.5 // 123.4	P = 0.241, ES: Small	13%
Napier et al., 2018	Prospective/VILR	Injured: 22, Uninjured: 33 (F)	15 weeks	Not reported	P > 0.05	Not reported
Gerlach et al., 2005	Prospective/VILR	Injured: 48, Uninjured: 39 (F)	1 year	117.4 (0.4) // 124.8 (5.8)	P = 0.46	6%
Kuhman et al., 2016	Prospective/VILR	Injured: 10, Uninjured: 9 (X)	1 season	100.3 (19.9) // 93.5 (30.3)	P = 0.65, ES: Small	-7%
Bredeweg et al., 2013 (sub-group)	Prospective/VILR	Injured: 11, Uninjured: 66 (M)	9 weeks	96.7 (30.75) // 127.0 (39.7)	P < 0.05*	31%
Davis et al., 2016 (sub-group)	Prospective/VILR	Injured: 11, Uninjured: 21 (F)	2 years	73.1 (15.9) // 88.0 (13.9)	P = 0.014*, ES: Large	21%
Ferber et al., 2002	Retrospective/VALR	TSF:10, Uninjured: 10 (F)	-	77.5 (29.4) // 117.9 (29.4)	P = 0.03*	52%
Ribeiro et al., 2015	Retrospective/VALR	CPF: 15, Uninjured: 30 (X)	Symptomatic	38.7 (9.7) // 64.4 (19.5)	P = 0.034*, ES: Large	39%
		APF: 30, Uninjured: 30 (X)	Symptomatic	38.7 (9.7) // 52.9 (13.9)	P = 0.001*, ES: Moderate	19%
Bigouette et al., 2016	Retrospective/VAL	CAI: 11, Uninjured: 13 (X)	Symptomatic	77.8 (10.0) // 93.8 (0.9)	P = 0.001*, ES: Moderate	21%
Milner et al., 2006	Retrospective/VALR	TSF: 20, Uninjured: 20 (F)	35 ± 28 months	66.3 (19.5) // 79.0 (25.0)	P = 0.041*, ES: Moderate	19%
Esculier et al., 2015	Retrospective/VALR	PPS: 21, Uninjured: 21 (X)	Symptomatic	69.7 (21.8) // 68.0 (17.3)	P = 0.78	-2%
Ferber et al., 2002	Retrospective/VILR	TSF: 10, Uninjured: 10 (F)	No access	108.9 (41.8) // 158.6 (41.8)	P = 0.03*	46%
Hreljac et al., 2000	Retrospective/VILR	Lower Limb: 20, Uninjured: 20 (X)	> 3 months	76.6 (19.5) // 93.1 (23.8)	P = 0.001*	22%
Pohl et al., 2009	Retrospective/VILR	PF: 25, Uninjured: 25 (F)	2.8±2.4 years	82.9 (18.7) // 100.5 (36.0)	P = 0.037*, ES: Small	21%
Milner et al., 2006	Retrospective/VILR	TSF: 20, Uninjured: 20 (F)	35±28 months	79.7 (18.8) // 92.6 (24.7)	P = 0.036*, ES: Moderate	16%
Messier et al., 1991	Retrospective/VILR	PPS: 16, Uninjured: 20 (X)	Symptomatic	53.9 (3.1) // 56.5 (4.5)	P > 0.05	5%
Azevedo et al., 2009	Retrospective/VILR	AT: 21, Uninjured: 21 (X)	Symptomatic	42.9 (9.3) // 44.8 (11.3)	P = 0.580	5%
McCrory et al., 1999	Retrospective/VILR	AT: 31, Uninjured: 58 (X)	Symptomatic	54.9 (1.7) // 55.5 (2.7)	P > 0.05	1%
Esculier et al., 2015	Retrospective/VILR	PPS: 21, Uninjured: 21 (X)	Symptomatic	83.1 (15.1) // 81.4 (15.0)	P = 0.716	-2%
Duffey et al., 2000	Retrospective/VILR	AKP: 99, Uninjured: 70 (X)	Symptomatic	54.9 (1.8) // 50.0 (1.7)	P < 0.05*	-9%

(F): female runners, (M): male runners, (X): mixed sex runners, TSF: tibial stress fracture, CPF: chronic plantar fasciitis, APF: acute plantar fasciitis, CAI: chronic ankle instability, PPS: patellofemoral pain syndrome, PF: plantar fasciitis, AT: achilles tendinitis, AKP: anterior knee pain, *: statistically significant.

With respect to vertical impact peak, 16 studies (4 prospective studies and 12 retrospective studies) [see table 2.3] have investigated the relationship with RRIs. Of the 4 prospective studies, one study Davis et al. [31] within a sub-group found 14% significantly higher vertical impact peak associated with the injured group compared to the uninjured (never injured) group. It is important to again mention that the larger study of vertical impact peak in Davis et al. [31] from which this sub-group was taken, found no significant relationship between vertical impact peak and RRIs; the sub-group which found the significant relationship specifically focused on those who had sustained their first injury versus those who had never been injured. This sub-group analysis was the only prospective study to find a significant relationship between vertical impact peak and RRI.

Regarding the retrospective studies, studies reported mixed directions when a significant relationship between vertical impact peak and a specific RRI was found. Specifically for chronic ankle instability [62] and lower limb injuries [77] examined in mixed sex groups, studies by Hreljac [62] and Bigouette et al. [77] reported higher magnitude of vertical impact peak in the injured group, with difference in magnitudes ranging from 13% to 21%. Conversely, Duffey et al. [28] analysed a relationship between anterior knee pain and vertical impact peak, also in a mixed group of male and female runners, and found a significantly lower magnitude of vertical impact peak was associated with injury.

Overall, it appears that there is moderate evidence to suggest a link between loading rates in terms of both VILR and VALR and risk of RRIs, with research suggesting that higher rates of loading (both VILR and VALR) are associated with RRIs, in particular for tibial stress fractures and plantar fasciitis. With regard to the vertical impact peak, there is little evidence to suggest a relationship with RRI.

Table 2.3: The relationship between vertical impact peak and running related injuries.

Author	Study Type	Population	Running Injury Time-Frame	Mean(SD) BW Uninjured//Injured	Significance (p value, Effect Size)	% Difference
Davis et al., 2016 (sub-group)	Prospective	Injured: 11, Uninjured: 21 (F)	2 years	1.51 (0.22) // 1.72 (0.21)	P = 0.013*, ES: Large	14%
Davis et al., 2016	Prospective	Injured: 144, Uninjured: 105; (F)	2 years	1.66 (0.31) // 1.67 (0.29)	P = 0.883, ES: Small	0.6%
Gerlach et al., 2005	Prospective	Injured: 48, Uninjured: 39 (F)	1 year	1.89 (0.05) // 1.87 (0.05)	P = 0.85	-1.1%
Bredeweg et al., 2013	Prospective	Injured: 34, Uninjured: 176 (X)	9 weeks	1.34 (0.16) // 1.29 (0.19)	P > 0.05	-3.7%
Messier et al., 2018	Prospective	Injured: 199, Uninjured: 101 (X)	2 years	1.55 (0.34) // 1.48 (0.34)	P = 0.74	-4.6%
Bigouette et al., 2016	Retrospective	CAI: 11, Uninjured: 13 (X)	Symptomatic	1.69 (0.20) // 2.05 (0.24)	P = 0.001*	21.0%
Hreljac et al., 2000	Retrospective	Lower Limb and Knee: 20, Uninjured: 20 (X)	> 3 months	2.13 (0.42) // 2.40 (0.41)	P < 0.05*	12.7%
Milner et al., 2006	Retrospective	TSF: 20, Uninjured: 20 (F)	35 ± 28 months	1.70 (0.32) // 1.84 (0.21)	P = 0.057	8.3%
Pohl et al., 2009	Retrospective	PF: 20, Uninjured: 20 (F)	2.8±2.4 years	1.70 (0.30) // 1.84 (0.30)	P = 0.093, ES: Moderate	8.3%
Azevedo et al., 2009	Retrospective	AT: 21, Uninjured: 21 (X)	Symptomatic	1.34 (0.20) // 1.45 (0.23)	P = 0.14	7.5%
McCrory et al., 1999	Retrospective	AT: 31, Uninjured: 58 (X)	Symptomatic	1.73 (0.04) // 1.81 (0.08)	P > 0.05	46%
Esculier et al., 2015	Retrospective	PPS: 21, Uninjured: 21 (X)	Symptomatic	3.1 (0.3) // 3.0 (0.4)	P = 0.339	-3.1%
Crossley et al., 1999	Retrospective	TSF: 23, Uninjured: 23 (M)	1.9±1.3 years	1.97 (0.34) // 1.89 (0.39)	P > 0.05	-4.1%
Messier et al., 1995	Retrospective	IBS: 56, Uninjured: 70 (X)	Symptomatic	1.75 (0.04) // 1.64 (0.04)	P > 0.05	-6.2%
Bennell et al., 2005	Retrospective	TSF: 13, Uninjured: 23 (X)	1.6±1.0 years	2.08(0.38) // 1.94 (0.30)	P = 0.32	-6.7%
Messier et al., 1991	Retrospective	PPS: 16, Uninjured: 20 (X)	Symptomatic	1.79(0.09) // 1.70 (0.09)	P > 0.05	-7.9%
Duffey et al., 2000	Retrospective	AKP: 99, Uninjured: 70 (X)	Symptomatic	1.74(0.04) // 1.66(0.31)	P < 0.05*	-4.6%

(F): female runners, (M): male runners, (X): mixed sex runners, TSF: tibial stress fracture, CAI: chronic ankle instability, PPS: patellofemoral pain syndrome, PF: plantar fasciitis, AT: achilles tendinopathy, AKP: anterior knee pain, IBS: iliotibial band syndrome, *: statistically significant.

The Relationship between Impact Accelerations and Injury

Even though studies have demonstrated a reasonable association between loading rates (VALR and VILR) and impact peak GRF [1], vGRF describes loading related to the whole body as opposed to segmentally. Therefore, it is possible that vGRF are not suitable to identify RRI risk factors described at a segmental level. The ability to measure loading at a segmental level has been achieved using (wireless or wired) wearable accelerometer sensors [27,79]. Due to their portability, low cost and light weight nature, wearable accelerometers have been incorporated in research focusing on RRIs to measure localised segmental loading.

It appears however that there are no prospective studies which investigate the relationship between RRI and peak impact accelerations. Retrospectively, 5 studies have examined the relationship with RRI, and 2 studies have reported that significantly greater peak impact accelerations were found in female runners who had experienced tibial stress fractures [27] and lower extremity stress fractures [32], compared to those female runners who had not. Two of the five studies additionally analysed peak tibial impact acceleration differences between the previously injured and uninjured lower limbs of runners who had previously experienced tibial stress fractures [80] and general RRIs [81]. It was found that in runners who had been previously injured, the magnitude of peak impact accelerations in their previously injured tibia was significantly greater in comparison to their uninjured tibia. Due to the retrospective nature of the findings, it is difficult to assess whether the injury had caused the differences, or whether the differences caused the injury.

Table 2.4: The relationship between (tibial) peak positive acceleration and running related injuries.

Author	Study Type	Population	Running Injury Time-Frame	Mean(SD) (g) Uninjured//Injured	Significance (p value, Effect Size)	% Difference
Milner et al., 2006	Retrospective	TSF: 20, Uninjured: 20 (F)	35 ± 28 months	5.8 (1.7) // 7.7 (3.2)	P = 0.014*, ES: Moderate	33%
Ferber et al., 2002	Retrospective	SF: 10, Uninjured: 10 (F)	Not reported	7.2 // 9.3	P = 0.05*	22%
Schutte et al., 2018	Retrospective	MTSS: 14, Uninjured: 16 (X)	Not reported	6.4 (1.5) // 6.6 (1.2)	p = 0.14, ES: Moderate	3%
Zifchock et al., 2006	Retrospective	TSF: 24, Uninjured: 25 (F)	Not reported	Not reported	p= 0.70	Not reported
Zifchock et al., 2006 (sub-group)	Retrospective	TSF: 24, Uninjured: 25 (F)	Not reported	Not reported	p =0.02*	16%
Zifchock et al., 2008	Retrospective	Injured: 20, Uninjured: 20 (X)	> 4 months	5.2 (2.4) // 5.1 (1.7)	p = 0.20, ES: small	-2%
Zifchock et al., 2008 (sub-group)	Retrospective	Injured: 20, Uninjured: 20 (X)	> 4 months	4.8 (1.6) // 5.5 (2.2)	p = 0.05*, ES: Moderate	15%

(F): female runners, (X): mixed sex runners, TSF: tibial stress fracture, SF: lower limb stress fracture, MTSS: medial tibial stress syndrome *: statistically significant.

No studies appear to have investigated the possible link between rate of impact accelerations and RRI, despite studies showing clearly that the loading rates of the vGRF were more related to RRI, than the vGRF impact peak [1]. As vGRF describes a measure of loading on the whole body, it is possible describing localised loading rate using a measure of rate of accelerations at those segments, may provide further understanding into how impact accelerations are related to RRI. Clearly there is a need for more research into the link between peak impact accelerations and RRI.

2.2.3.2   Technique Factors that Affect Injury

While joints and body segments are capable of rotations about three axes, this review focuses solely on movements in the sagittal plane, as these have the largest angles of excursion, and contribute most to the impact accelerations and the vGRFs, reviewed above.

The Relationship between Foot Strike Pattern and Injury

Foot strike patterns can be classified as either: rear-foot strike, mid-foot strike or fore-foot strike [82]. Mid-foot and fore-foot strike patterns are often referred to as non-rear-foot strike patterns. Recent research has been interested in how different foot strike patterns relate to injury [29,83,84]. Kinematics and kinetics during running can be influenced by foot strike pattern due to the differences in how initial contact is made with the ground (see Section 2.1, [29]). A large-scale study of 514 recreational shod runners by De Almeida et al. found that just over 95% of runners in the study were identified as have a rear-foot strike [85].

With respect to kinematics, rear-foot runners tend to contact the ground with the knee slightly flexed, in a way which puts pressure on the knee muscles to absorb the impact shock-wave from the GRF. This is different to the fore-foot strike pattern which usually has a more flexed knee position at initial contact [86]. Kinematics of a mid-foot strike pattern are transitional between fore-foot and rear-foot strike patterns [82].

In examining the relationship between foot strike technique and RRIs, the foot strike technique has been looked at in terms of a categorical measure of foot strike pattern and a continuous measure of foot strike technique. Where categorical measures refer to rear-foot strike, mid-foot strike or fore-foot strike [82] and continuous refer to measure such as foot contact angle, strike index and ankle flexion angle. Foot contact angle is the angle of the foot relative to the ground at initial contact [68]. Strike index specifies where the centre of pressure is located in regard to the long axis of the foot at initial contact [87]. Ankle flexion angle describes the flexion of the ankle measured at initial contact.

Twelve studies appear to have explored the relationship between RRIs and foot strike technique. Five of these studies examined categorical foot strike pattern, while the remaining seven studies examined continuous measures of foot strike technique. All five categorical studies were retrospective in design, no prospective studies were found. Three of the five studies found that a rear foot strike pattern was more significantly associated with injured runners [29, 82, 88]. One of the five studies, found that non-rear foot strike patterns were approximately 2 times more likely to sustain a specific RRI [30]. The remaining study by Warr et al. [89] reported no significant findings.

With regard to studies examining continuous measures of foot strike technique, three of the seven studies were prospectively designed, and the remaining four studies were retrospective. Prospectively, none of the three studies (Dudley et al. [68], Messier et al. [90] and Kuhman et al. [71]) reported significant findings while one of the retrospective studies (Dingenen et al. [91]) found a significant relationship between foot contact angle and RRIs, with on average approximately 3 degree smaller foot contact angles found in injured limbs. The three remaining retrospective studies by Mann et al. [92], Donoghue et al. [93], and Paquette et al. [83] report no significant findings.

There appears to be relatively low evidence suggesting an association between foot strike technique overall and RRIs. However, there appears to be greater evidence of relationship between categorical foot strike technique and RRIs compared to continuous measures. Four of five studies which examined categorical measures found a significant relationship with RRIs. Due to mainly retrospective studies with small population sizes, it is difficult to be sure of the relationships reported. Therefore, there is a need for future prospective research to consider foot strike technique in relation to RRIs.

The Relationship between Knee Kinematics and Injury

Research has highlighted the importance of knee kinematics at initial contact to absorb the shock-wave associated with ground impact during foot strike [94]. There is speculation on the effect of knee

kinematics on landing strategy at initial contact and how this is related to RRI [95–97]. Regarding knee kinematics in the sagittal plane, knee flexion at initial contact and peak knee flexion have been notable areas of research interest. Knee flexion angles are calculated from full extension, with a straightened leg at zero degrees, as flexion angle increases with flexion and decreases with extension.

Few studies have investigated knee flexion at initial contact in relation to injury, with one prospective study [98] and two retrospective studies [33, 99] found. Only one of the studies reported a significant finding, where Bramah et al. [33] found significantly smaller knee flexion angles (an absolute difference of 4 degrees) at initial contact were associated with a greater risk of injury. The smaller knee flexion angles suggest landing in a pose which is more extended than flexed. This may cause greater vGRFs at impact and less attenuation of the forces throughout the lower limbs. [94, 100, 101].

Table 2.5: The relationship between knee flexion at initial contact and running related injuries.

Author	Study Type	Population	Running Injury Time-Frame	Mean(SD) (Degrees) Uninjured//Injured	Significance (p value, Effect Size)	Absolute Difference
Noehren et al., 2007	Prospective	IBS:18, Uninjured: 18 (F)	2 years	14.4 (6.0) // 11.8 (4.8)	P = 0.18	-3 °
Milner et al., 2007	Retrospective	TSF: 23, Control: 23 (F)	Not reported	11.9 (6.5) // 13.7 (6.0)	P = 0.35, ES: Small	2°
Bramah et al., 2018	Retrospective	Injured: 72, Uninjured: 36 (X)	Symptomatic	10.2 (4.8) // 6.0 (4.9)	P < 0.01*, ES: Large	-4°

(F): female runners, (M): male runners, (X): mixed sex runners, TSF: tibial stress fracture, IBS: iliotibial band syndrome, *: statistically significant.

With the number of studies examining knee flexion at initial contact rather low, the relationship between RRIs and peak knee flexion can be additionally explored, as it has been the focus of more research. Eight studies (2 prospective and 6 retrospective) were identified. Only one of the studies reporting a significant finding; retrospectively Loundon et al. [102] found that significantly less peak knee flexion angles (an absolute difference of 5 degrees) were associated with the injured group of runners when compared to the peak knee flexion of the uninjured group. See section 2.2.3.3 for an examination of the relationship between knee kinematics and loading.

Table 2.6: The relationship between peak knee flexion and running related injuries.

Author	Study Type	Population	Running Injury Time-Frame	Mean(SD) (Degrees) Uninjured//Injured	Significance (p value, Effect Size)	Absolute Difference
Messier et al., 2018	Prospective	Injured: 199, Uninjured: 101 (X)	2 years	40.1 (4.7) // 40.0 (5.3)	P = 0.82	-0.1°
Hein et al., 2014	Prospective	AT:10, Uninjured: 10 (X)	1 year	41.0 (4.0) / 37.0 (7.0)	Not reported	-4°
Wirtz et al., 2012	Retrospective	PPS: 20, Uninjured: 20 (F)	Symptomatic	41.8 (4.1) /43.9 (5.0)	P = 0.16, ES: Moderate	2°
Ferber et al., 2010	Retrospective	IBS: 35, Uninjured: 35 (F)	N/R	45.2 (5.0) // 45.3 (4.5)	P = 0.95	0.1°
Bramah et al., 2018	Retrospective	Injured: 72, Uninjured: 36 (X)	Symptomatic	10.2 (4.8) // 6.0 (4.9)	P < 0.01*, ES: Large	-4°
Azevedo et al., 2009	Retrospective	AT: 21, Uninjured: 21 (X)	Symptomatic	42.8 (8.6) //42.2 (4.8)	P = 0.80	-0.6°
Loudon et al., 2012	Retrospective	MSP: 14, Uninjured: 14 (X)	2 years	42.1 (4.8) // 37.1 (5.4)	P = 0.02*	-5°
Bazett-Jones et al., 2013	Retrospective	PPS:19, Uninjured: 19 (X)	Symptomatic	54.1 (4.8) // 41.6 (5.4)	P > 0.05	-13°

(F): female runners, (M): male runners, (X): mixed sex runners, PPS: patellofemoral pain syndrome, AT: achilles tendinopathy, IBS: iliotibial band syndrome, MSP: medial shin pain, *: statistically significant.

Overall, the findings in relation to knee kinematics and injury are mixed, with a lack of prospective studies examining both peak knee flexion and knee flexion at initial contact with respect to injury.

Although as noted above, some studies reported a significant finding concerning knee flexion at initial contact [33], and peak knee flexion [102], there are not enough studies or evidence to conclusively determine if there is a link between these knee kinematics and RRIs.

The Relationship between Hip Kinematics and Injury

Tiberio [103] suggests that hip movements may be due to other movements in the lower limbs which have a knock-on effect, e.g., exaggerated foot motion may impact hip kinematics. These changes in kinematics may result in injury, and so researchers have explored a range of areas within hip kinematics pertaining to injury with much focus on peak hip adduction [68, 104] and peak hip internal rotation [68, 72]. Peak hip flexion and hip flexion excursion have been overlooked in terms of their potential importance in relation to RRI research. While no prospective studies could be found examining peak hip flexion, retrospective studies have analysed the differences in peak hip flexion angles between an injured and injured mixed sex group of runners, with specific interest in patellofemoral pain syndrome in Bazett-Jones et al. [105] and iliotibial band syndrome in Grau et al. [106]. Neither study found any significant relationship.

Table 2.7: The relationship between peak hip flexion and running related injuries.

Author	Study Type	Population	Running Injury Time-Frame	Mean(SD) (Degrees) Uninjured//Injured	Significance (p value, Effect Size)	Absolute Difference
Grau et al., 2011	Retrospective	IBS: 18, Uninjured: 18 (X)	Symptomatic	32.0 (6.0) // 31.0 (4.0)	P > 0.05	-1.0°
Bazett-Jones et al., 2013	Retrospective	PPS: 19, Uninjured: 19 (X)	Symptomatic	35.8 (8.4) // 30.4 (6.8)	P = 0.73	-5.4°

(X): mixed sex runners, PPS: patellofemoral pain syndrome, IBS: iliotibial band syndrome, *: statistically significant.

No prospective studies could be found examining hip flexion excursion. One retrospective study by Grau et al. [106] examined the differences in hip flexion excursion angles between an injured and uninjured mixed sex group of runners, with specific interest in Iliotibial band syndrome. Grau et al. found no significant findings in relation to hip flexion excursion. Due to the lack of evidence, and prospective studies, it is not possible to conclude on an association between running injuries and hip kinematics.

Table 2.8: The relationship between hip flexion excursion and running related injuries.

Author	Study Type	Population	Running Injury Time-Frame	Mean(SD) (Degrees) Uninjured//Injured	Significance (p value, Effect Size)	Absolute Difference
Grau et al., 2011	Retrospective	IBS: 18, Uninjured: 18 (X)	Symptomatic	45.0 (5.0) // 44.0 (3.0)	P > 0.05	-1.0°

(X): mixed sex runners, IBS: iliotibial band syndrome, *: statistically significant.

The Relationship between Trunk Kinematics and Injury

To date, trunk kinematics in relation to RRI have not been extensively researched. With respect to trunk kinematics in the sagittal plane, it would appear that no prospective studies, and a single retrospective study (Bazett et al. [105]) have investigated peak trunk flexion with regard to injury. Bazett et al. explored the differences in peak trunk flexion angles between an injured and injured mixed sex group of runners, with specific interest in patellofemoral pain syndrome, and found no significant findings. Due to the lack of research or significant findings it is not possible to conclude if there is any relationship between trunk kinematics and running injury.

Table 2.9: The relationship between peak trunk flexion and running related injuries.

Author	Study Type	Population	Running Injury Time-Frame	Mean(SD) (Degrees) Uninjured//Injured	Significance (p value, Effect Size)	Absolute Difference
Bazett-Jones et al., 2013	Retrospective	PPS: 19, Uninjured: 19 (X)	Symptomatic	13.9 (4.7) // 13.1 (6.2)	P = 0.88	-0.8°

(X): mixed sex runners, PPS: patellofemoral pain syndrome, *: statistically significant.

2.2.3.3 Technique Factors that Affect Loading

In light of the limited research examining the relationship between running technique and injury, this section describes how technique affects impact loading because of the proposed relationship between increased loading and injury [27]. First, foot strike patterns are discussed and the affect that differing foot strike patterns have on measures of loading (VALR, VILR, impact accelerations and vertical impact peak). Following this is a description on how knee angles at initial contact affect loading on the body. Finally, various running technique styles (Pose and Chi) which aim to adapt running kinematics and loading are discussed.

The Relationship between Foot Strike Kinematics and Loading

As previously mentioned, there are three different types of foot strike patterns: rear-foot strike, mid-foot strike and fore-foot strike [82]. Mid-foot and fore-foot strike patterns are often referred to as non rear-foot strike patterns also.

A number of studies have examined the relationship between foot strike and loading, using a variety of different measures of loading (VIP, VALR, VILR and acceleration). Regarding vertical impact peak, there are mixed findings in relation to the effect of rear-foot versus fore foot running. Thompson et al. [107], Mercer et al. [108] and Kulmala et al. [86] reported that significantly larger vertical impact peaks were found in the rear-foot running pattern, whereas Kuhman et al. [109] and

Laughton et al. [110] found the opposite result, that vertical impact peak was significantly lower during the rear-foot running strike pattern. In addition, Goss and Gross [111] found no relationship between vertical impact peak and varied running strike patterns.

Table 2.10: The relationship between vertical impact peak and foot strike pattern.

Author	Population	Foot Strike Patterns	Findings	Significance (p value, Effect Size)
Kulmala et al., 2013	38 runners (F)	RFSP vs FFSP	VIP 26% greater in RFSP	P = 0.001*, ES: 2.43 Large
Mercer et al., 201	10 runners (M)	RFSP vs MFSP vs FFSP	VIP greater in RFSP than FFSP	P < 0.05*
Thompson et al., 2015	10 runners (X)	RFSP vs NSP	VIP 15% greater in RFSP	P < 0.05*
Goss et al., 2013	44 runners (X)	RFSP vs Chi Running	-	P = 0.61
Kuhman et al., 2015	16 runners (M)	RFSP vs FFSP	Peak vGRF 5.2% lower in RFSP	P < 0.001*, ES: 0.65 Small
Laughton et al., 2003	15 runners (Un)	RFSP vs FFSP	Lower peak vGRF in RFSP	P = 0.0002*

(F): female runners, (M): male runners, (X): mixed sex runners, (Un): unknown sex runners, RFSP: rear-foot strike pattern, MFSP: mid-foot strike pattern, FFSP: fore-foot strike pattern, NSP: natural strike pattern, *: statistically significant.

With respect to rates of loading (VALR and VILR), there are strong evidence to suggest a link between loading rates and differing foot strike patterns. The majority of studies have reported significantly lower VALRs (37%-47%) and VILRs were associated with non-rear foot strike runners (fore foot runners [86, 109, 110, 112, 113], and chi runners [111]) in comparison to rear-foot strike runners (see Chi running in Section 2.2.3.3).

Table 2.11: The relationship between loading rates (VALR and VILR) and foot strike pattern.

Author	Loading Rate	Population	Foot Strike Patterns	Findings	Significance (p value, Effect Size)
Kulmala et al., 2013	VALR	38 runners (F)	RFSP vs FFSP	VALR 47% lower in FFSP	P = 0.001*, ES: 2.44 Large
Kuhman et al., 2015	VALR	16 runners (M)	RFSP vs FFSP	VALR greater in RFSP	P < 0.001*, ES: 2.09 Large
Goss et al., 2013	VALR	44 runners (X)	RFSP vs Chi	VALR 37% lower in Chi runners	P < 0.001*
Shih et al., 2013	VALR	12 runners (M)	RFSP vs FFSP	VALR greater in RFSP	P < 0.000*
	VILR	12 runners (M)	RFSP vs FFSP	VILR greater in RFSP	P < 0.000*
Laughton et al., 2003	VALR	15 runners (Un)	RFSP vs FFSP	-	P = 0.99
Yong et al., 2018	VALR	17 runners (X)	RFSP vs FFSP	VALR greater in RFSP	P < 0.05*
	VILR	17 runners (X)	RFSP vs FFSP	VILR greater in RFSP	P < 0.05*

(F): female runners, (M): male runners, (X): mixed sex runners, (Un): unknown sex runners, RFSP: rear-foot strike pattern, FFSP: fore-foot strike pattern, Chi: chi running pattern *: statistically significant.

The effect of differing running foot strike pattern on impact accelerations has not been extensively examined to date, as it appears to only feature in four studies investigating peak positive acceleration. Of these four studies, three reported significant findings in conflicting directions. Ruder et al. [114] and Delgado et al. [115] found higher peak positive acceleration associated with rear-foot strike runners in comparison to fore foot strike runners, while Laughton et al. [110] found lower peak positive accelerations were associated with a rear-foot strike pattern. Interestingly, Young et al. [112] found no significant differences between the two foot strike patterns.

The strongest evidence of a relationship can be seen between loading rates (VILR and VALR) and differing foot strike patterns, with the majority of studies relating higher rates of loading with

rear-foot runners. It is still unclear if there is a relationship between impact acceleration and differing foot strike pattern due to the lack of research and mixed findings within the studies which exist. This can also be said for vertical impact peak and foot strike pattern, as there were mixed findings among studies, which leaves the relationship between them inconclusive.

Table 2.12: The relationship between impact acceleration and foot strike pattern.

Author	Population	Foot Strike Patterns	Analysed	Findings	Significance (p value)
Laughton et al., 2003	15 runners (Un)	RFSP vs FFSP	TPPA	Lower TPPA in RFSP	P = 0.03*
Yong et al., 2018	17 runners (X)	RFSP vs FFSP	TPPA	-	P > 0.05
Delgado et al., 2013	43 runners (X)	RFSP vs FFSP	TPPA	Greater TPPA in RFSP	P < 0.001*
Ruder et al., 2017	222 runners (X)	RFSP vs MFSP vs FFSP	TPPA	Greater TPPA in RFSP and MFSP than FFSP	P = 0.01*

(X): mixed sex runners, (Un): unknown sex runners, RFSP: rear-foot strike pattern, MFSP: mid-foot strike pattern, FFSP: fore-foot strike pattern, TPPA: tibial peak positive acceleration, *: statistically significant.

The Relationship between Knee Kinematics and Loading

The effect of changes in knee angle on vGRFs during running has not been extensively examined. However, some studies have observed an inverse relationship, with decreases in vGRF associated with increases in knee joint flexion at initial contact during running [116]. However, these were secondary to other changes: fore-foot versus rear-foot [117], use of orthotics [110], and shod versus barefoot [117] and adoption of a compliant running technique [26]. In addition, in a mathematical simulation study of running Gerritsen et al. [116] demonstrated for every 1° increase in knee flexion there is a subsequent 68N decrease in vertical impact peak.

In light of the lack of direct evidence examining the effect of changes in knee angle and vGRF during the landing phase of running, it is worth considering other tasks where changing knee angle has been examined. For example, the inverse relationship has also found in experimental studies [118,119] and modelling of the landing phase of the drop jump [120]. The inverse relationship is also evident when using a horizontal pendulum to swing participants towards a wall mounted force plate, which allows for an even greater control of knee angle [121,122]. Potthast et al. [122] showed that as knee angle changed from 0 degrees (fully extended) to 40 degrees, the force associated with the vertical impact peak was reduced by approximately 160N.

The inverse relationship between vGRF and knee angle can be explained in part by the resultant GRF vector being orientated away from the knee and hip joint centres (posterior to the knee and anterior to the hip). This allows the body's active structures, the muscle-tendons, to absorb and attenuate the GRF, predominantly through eccentric contraction [123]. In contrast, when the knee joint is in a more extended position, the GRFs are directed more through the skeleton and the passive

structures become the more dominant attenuating structures, resulting in less load attenuation [116].

The Relationship between Running Styles and Loading

Various running styles which aim to adapt running kinematics have been developed, such as Pose running [124] and Chi running [125]. While Pose running and Chi running appear to have gained some popularity amongst recreational runners, neither of these two running styles have been researched to any significant extent. However, given that those who advocate their use suggest they can reduce loading and reduce the likelihood of injury [126], it is important to briefly review them.

The Pose running technique is defined by striking the ground with a mid-to fore-foot strike pattern, while keeping ground contact time short [124, 127], and maintaining a forward leaning position with consistent flexion of the knee [124, 128]. It appears to only feature in three studies (Arendse et al. [129], Fletcher et al. [130] and Wei et al. [131]) with mixed findings. Arendse et al. [129] compared the rate of loading and vertical impact peaks associated with rear-foot strike patterns and Pose running in recreational runners and found that rear-foot running patterns resulted in significantly higher rates of loading and impact peaks (both approximately 40%) in comparison to the Pose running technique. However, no significant findings were found in the remaining two studies which also tested on recreational runners.

Chi running is characterised by 'a mentality aligning the body and mind', in a forward leaning position, and avoiding heel strikes associated with rear-foot strike patterns [111]. The effect of Chi running on loading appears to only be investigated in one study by Goss and Gross [111]. Goss and Gross compared VALR of Chi runners and rear-foot strike runners and reported that larger VALRs were associated with rear-foot runners.

Due to the lack of research, it is not possible to conclude if there is any relationship between differing running styles and loading on the body. There is no strong evidence suggesting Pose running or Chi running styles are beneficial in reducing RRIs.

2.2.4 Non-Biomechanical Risk Factors Associated with Running-Related Injuries

Other than loading and technique (reviewed in sections above) there are other intrinsic risk factors (sex, age, running experience and previous injury) and extrinsic risk factors (training load, surface) for running related injuries. These factors are reviewed in brief below.

2.2.4.1 Intrinsic Risk Factors

Sex

With regard to sex, prospective studies found little evidence of a relationship with RRIs, and retrospective studies found only moderate supporting evidence. Some studies examined general RRI risk and others explored risk related to specific injuries. With regard to general running related injuries the findings appear to be mixed. For example, a prospective study by Buist et al. [132] with 629 recreational runners found that male runners were at a statistically greater risk of general running related injuries (hazard ratio: 1.42), while a prospective study by Messier et al. [90] which involved 300 recreational runners and found that female runners were at a statistically greater risk of general RRIs (hazard ratio: 1.42). Similar conflicting results were found for general RRIs retrospectively with Lopes et al. [133] reporting a statistically greater risk of general RRIs for female runners, and Linton et al. [134] reporting a statistically greater risk of general RRIs for male runners, where each study had >1000 participants.

Studies which focused on specific RRIs found that the risk of specific injuries may be greater due to sex. A prospective study by Satterthwaite et al. [135] with 875 recreational runners found a statistically greater risk of hip injury in female runners (odds ratio:1.88, p=0.01), and a statistically greater risk of calf injury (odds ratio:1.86, p=0.0008) and hamstring injury (OR:1.60, p=0.03) in male runners. This suggests that future studies should take sex into account as a risk factor for specific RRIs rather than general RRIs as a whole.

Age

With regard to age, prospective and retrospective studies have found moderate proof of a relationship with RRIs, although the reporting on the direction of this relationship is mixed. Some studies examined general RRI risk and others explored risk related to specific injuries. Two prospective studies which focused on general RRIs found that higher age had a significantly greater risk of RRI [136, 137]. In contrast a prospective study by Buist et al. [132] (prospective) and retrospective studies Hootman et al. [138] and Rasmussen et al. [139] found that a lower age had a significantly greater risk of RRI.

Where the risk factor of specific injuries was taken into account, two prospective studies found that higher age was associated with a significantly greater risk of thigh injuries in Satterthwaite et al. [135] and achilles tendinopathies in Hirschmuller et al. [140]. However, Satterthwaite et al. [135] and

Kelsey et al. [141] prospectively found that lower age was associated with a significantly greater risk for calf injuries and stress fractures respectively. A moderate level of evidence across these studies suggests that there is a relationship between age and RRI risk, but the direction of this relationship is unclear.

It could be suggested that older runners maybe be more susceptible to injury due to the changes experienced by the musculoskeletal system with age, e.g., decreased strength and flexibility [142], as some studies have found increased age to be a greater risk for RRI [136]. Despite this, a number of studies have also reported that lower age is associated with a significantly greater risk of RRI as mentioned above. It is suggested that older runners with more running experience, and more training experience may be less likely to repeat training errors which may lead to injury, unlike new novice runners [143]. The lack of information on additional related factors i.e., previous injuries or running experience may have an impact on the mixed findings [139, 144].

Previous Injury

Previous injury seems to be an important risk factor for successive RRIs. Many studies have found a significantly greater risk of re-injury associated with having had a previous running injury (prospective studies [34, 132, 145, 146] and retrospective studies [134, 138, 139]). A number of reasons have been suggested for the strong association between previous injury and future RRIs. This may be due to previous injuries resulting in long term biomechanical traits that are harmful, leading to an increased risk for further injury. Runners recovering from injury may change their technique to compensate for the injured limb in order to shift the strain away from the site of injury [147], potentially causing excessive loading on joints leading to re-injury or new RRIs [148]. An additional suggestion is that past injuries may have not healed completely before returning to training [58]. In addition, if injury rehabilitation is inadequate for some reason i.e., poor adherence to rehabilitation programs, then re-occurrence of injury is likely to occur as strength of the injured limb may not have recovered before returning to training [149].

While most studies use data from the year previous to the study to define the time-frame for 'previous injuries' [134, 138, 139], some studies do not report the time between previous and subsequent injury, or the time-frame used to examine previous injuries [132, 140]. This makes it difficult to get a sense of how old a previous injury needs to be in order to remain a risk factor for RRIs.

Running Experience

In relation to running experience, many prospective studies found no significant difference in RRI risk [90, 141, 145, 150, 151]. Studies which found a significant relationship between running experience and risk of RRI report contrasting results. Wen et al. [34] prospectively and Hootman et al. [138] retrospectively found that increased running experience was related to a significantly greater risk of RRIs. In direct contrast, prospectively Kluitenberg et al. [137] and Buist et al. [132] found that lower running experience was related to a significantly greater risk of general RRIs.

The definition of running experience varies across studies, with some studies defining running experience with reference to number of months, or years of experience [152] and others defining experience based on races or events completed [138]. Past experience with injury, experience in other running related sports, and age can contribute to defining this factor, which adds to the complexity surrounding the definition of running experience [58]. Studies must consider the inclusion of potential confounding factors in the analysis of running experience as a risk factor for RRIs [134, 139].

2.2.4.2 Extrinsic Risk Factors

Surface

Running on different surfaces has been suggested to result in differing joint kinematics, muscle activity as well as impact loading [153]. There are a range of surfaces which may be considered for running, road, grass, sand, dirt treadmills, depending on the runner's location, accessibility and their personal preference. Studies provide little evidence to suggest that differing surfaces may affect the risk of RRI. Few studies have reported significantly greater risk of specific injuries when running on one surface versus another [35, 154, 155]. Many studies have reported no relationship between RRIs and differing running surfaces [34, 136, 141]. It has been suggested the lack of evidence of a relationship is due to the ability of the runner to change their technique across different surfaces e.g., changing knee angle at initial contact, in order to keep a similar level of impact loading across surfaces [94].

Training Load

Training load is the cumulative amount of stress placed on an individual from a single or multiple training sessions over a period of time [57]. Training load can be described as a combination of factors:

distance, duration, frequency and speed of training.

There is little evidence to support a risk associated with training distance and RRI. Significant findings suggest that further training distance leads to a significantly greater risk for general RRI (prospectively Ferreira et al. [156] and retrospectively Knobloch et al. [35]) and specifically hamstring injuries (prospectively Wen et al. [34] and Satterthwaite et al. [135]). However, conflicting results found prospectively in Di Caprio et al. [154] and Satterthwaite et al. [135] suggest that lower running distances were associated with higher risk of hamstring and knee injuries respectively. Many prospective studies found no relationship between training distance and risk of RRIs [140, 141, 145]. There are differences in how 'distance' was defined between studies with some reporting the absolute distance over days, weeks and months. Recent models have suggested that distance needs to be explored as a measure that is relative to the runner's usual distances as opposed to absolute distance [157]. Therefore, if the relative training load for a runner increases greatly in the short term, they may be more at risk to RRIs as it is excessive relative to their usual training load.

Similar to training distance, there is little evidence supporting an association between training duration and the risk of RRIs. Studies report mixed results, with prospective studies finding a significantly higher [158] and significantly lower [137] risk of injury associated longer training duration. Some prospective studies found no connection between training durations and risk of injury [140, 156, 159]. It is possible that other factors relating to training load have a joint impact with training duration on the likelihood of injury, but the analysis of running duration alone provides little insight.

There is little evidence to suggest training session frequency is a risk factor for RRI. Unlike previous training load factors (distance and duration), the significant findings are less mixed in direction, with the majority of studies (which find a significant relationship) suggesting that increased training frequency may result in an increased risk of incurring RRIs (prospectively [135, 154] and retrospectively [35, 160]). Increased frequency of training may cause a decrease in a runner's recovery time between sessions, which may increase the likelihood of injury due to strain and overuse [57]. The type of training session and the intensity associated with sessions could have an effect on whether higher frequency of sessions would be more demanding on a runner and their need for recovery time. These additional parameters should be considered along with frequency and other training load factors to understand how these impact risk of injury.

There are not many studies which have investigated training speed in relation to risk of injury, of

which there is little reported evidence, with many prospective studies finding no significant relationship [90, 159, 161]. Studies which reported a relationship consistently found that increased running speed was associated with a greater risk of becoming injured (Junier et al. [158] [prospective], Hootman et al. [138] and McCrory et al. [74] [retrospective]). Related to training speed is the effect of interval training, which has been investigated by an even smaller number of studies. Of the studies which have explored interval training and injury, there is little reported evidence of a relationship. Prospectively, Junior et al. [158] reported that performing more interval training significantly decreased the risk of RRI. Whereas in a separate prospective study, Wen et al. [34] found that runners who performed more interval training were at a significantly higher risk of shin injuries specifically. Due to the lack of studies, it cannot be definitively concluded that there is a relationship between either speed or interval training and risk of RRIs.

Chapter 3

Related Work: Orientation Estimation

3.1 Introduction

This section outlines various methods used for calculating orientation using inertial sensor data. As mentioned in the motivation of this thesis (Section 1), the key objective of this review is to identify the most accurate methods of calculating human motion orientation data, specifically for running. The accuracy of methods were specifically analysed in terms of; the reported segment orientation or joint angle accuracy, which is closely related to the approximate duration of data capture, and the movement examined. A number of factors appear to have significant influence on determining accuracy, and therefore will be considered within this review.

The methods can be divided into four groups: Gyroscope Integration, Stochastic methods, Complementary methods and Deterministic methods. The Stochastic methods described in Section 3.4, and the Complementary methods described in Section 3.5, appear to be the most used and most accurate approaches for human movement analysis [14, 162].

The type of movements that have been used for experimental testing can been classified in two ways. The first classification is based on the level of impacts present in the movement, where studies can be separated into three categories: experiments involving no impact (e.g., simple rotation movements performed around one or more axes) [163], experiments involving low impact (e.g., walking) [164], and experiments involving high impact (e.g., running, jumping) [8]. The second classification is based on movement speed and movement complexity, where studies can be separated into three categories: experiments involving slow speed and low movement complexity (e.g., simple rotation movements performed around one or more axes) [163], experiments involving moderate speed and moderate movement complexity (e.g., walking) [164], and experiments involving high speed and high movement complexity (e.g., running, cycling) [8]. It was required to review impact level and speed/complexity of movements separately in order to capture the description of all

movements clearly e.g., cycling, which is a high speed and high complexity movement which does not involve impacts. All testing movements used in the studies reviewed will be commented on based on impact levels of the movement, speed and complexity of the movement and duration of testing. References to pitch, roll and yaw angles refer to the angles of rotation calculated around the X, Y and Z axes, respectively, of an inertial sensor.

It is not within the scope of this thesis to provide an in-depth analysis of the mathematics underpining each of the methods, however high-level descriptions of each of the methods are provided. The aim of this review is to evaluate the most accurate and most appropriate methods of calculating orientation for running. The studies included in this review were chosen on the basis that they compared their experimental results to reliable reference data and quantified the magnitude of their deviation/error from the reference data.

3.2 Gyroscope Integration

The traditional method of calculating orientation involves numerically integrating angular velocity. Using this method to calculate orientation with sensor data relies solely on the gyroscope signal. Problems arise using this method as the noise and biases which exist in the gyroscope signals are included in the integration process, and therefore accumulate in the output orientation to become substantial errors [6, 12]. Over time, this leads to inaccurate orientation results if the gyroscope bias and noise errors are not accounted for [12, 165–167]. This can reduce the motion capturing window for accurate sensor data to less than one minute [167]. After one minute, these errors can grow by 10-25 degrees per second [168].

Computationally this is a mathematically easy method to apply, but given its increasing inaccuracy with time, it is not appropriate to use this method for measuring segment orientation or joint angles for human movement analysis applications. This is especially true for dynamic applications involving impacts e.g., running, as the presence of impacts may introduce additional noise to the data [169]. Additionally, given a key outcome of this literature review is to identify methods which can be used accurately for more than 30 minutes of data capture, this approach is not suited to the aims of this thesis. For this reason, this method was not considered for the detailed comparative section of this review (Section 3.6). Studies which apply other orientation estimation approaches (outlined in the following sections) often include a comparison of the results of their proposed method to results using gyroscope integration. This highlights the fact that this approach is outperformed by other approaches

to calculate orientation [12, 165].

3.3   Sensor Fusion

A solution to the issues caused by relying on the performance of one sensor is sensor fusion. Sensor fusion describes fusing information from multiple sensors under one framework to provide more meaningful combined information. In this context, sensor fusion is used to calculate orientation more accurately than if it were calculated using the information provided by one sensor alone. Orientation data must be reliable in order to calculate accurate segment orientations, joint angles, and position. The following sections provide an analysis of some of the current literature on this topic.

3.4   Stochastic Methods

Introduction

Stochastic methods appear to be a very common approach to calculate orientation [14,162] and Kalman filtering has been well studied in the area of human movement. It has been shown that they are more accurate than the gyroscope integration method [6, 12]. The first section below reviews the Linear Kalman Filter and is followed by sections describing two of its non-linear variants, the Extended Kalman Filter, and the Unscented Kalman Filter.

3.4.1   Linear Kalman Filter

The Kalman Filter was originally developed by Kalman in 1960 [170] and has since been used in a number of domain areas for estimation problems. Specifically, it has been a very popular estimation method for calculating orientation. The Kalman filter is a two-step estimation problem composed of a prediction step and an update step (figure 3.1). This approach iteratively predicts an estimate of the state of a linear system and predicts the uncertainty associated with the state estimate. In the update step, the model reads in actual measurement data i.e., the incoming measured noisy sensor data. The Kalman Gain is calculated using the uncertainty in the accuracy of current estimate and the uncertainty in the accuracy of the measured data. The algorithm looks at the difference between the current state estimate and the current measurement data and uses that result along with the Kalman gain to update the current estimate. Depending on the value of the Kalman gain, the current estimate will be updated with either more confidence given to the measurement value or the estimated

value. Each new set of measurement data leads to better estimation. Even when using a poor starting estimation of the state, the model will still converge towards the correct prediction of the future states. These two core steps, predict and update, are repeated recursively to minimise the error between the predicted current state and actual measurement values for the system. In its linear form with the error modelled as Gaussian noise, the Kalman filter is an optimal estimator. Full equations describing the Kalman filtering approach can be found in [171].
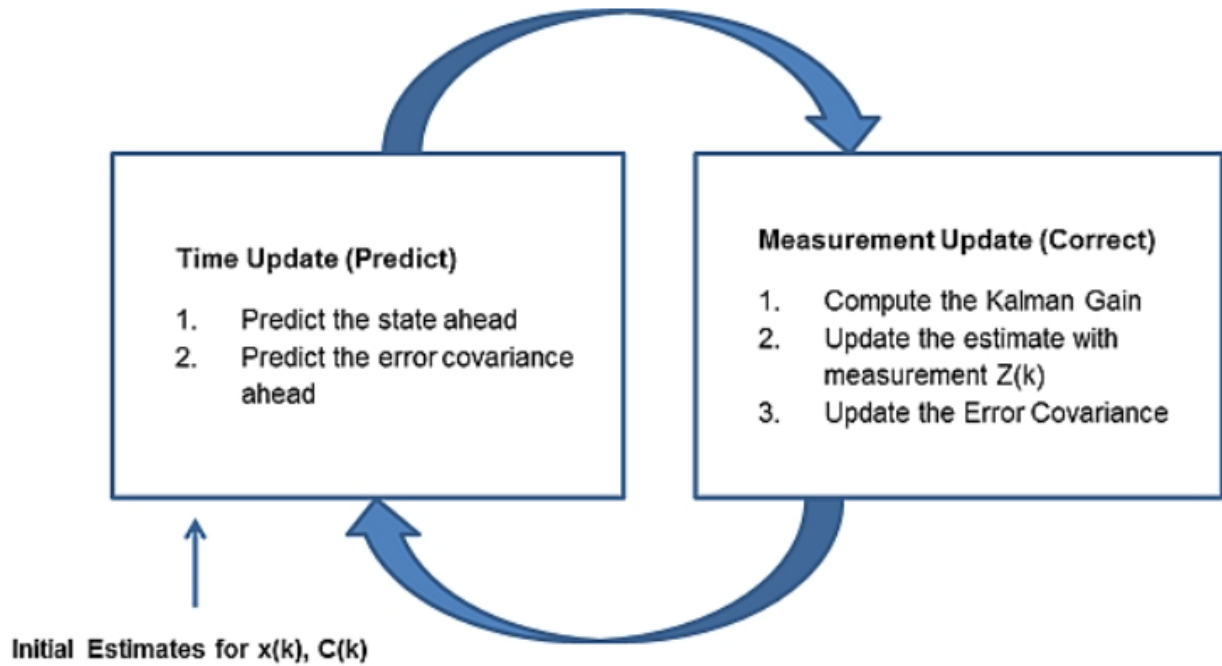
Figure 3.1: Steps involved in the Linear Kalman Filter.

Twenty studies (see studies in tables 3.1, 3.2 and 3.3) have been included in this review for the purpose of providing a scope of the area of Kalman Filtering. These 20 studies examined different variations of Kalman filtering (linear, extended, and unscented filtering) as an approach to calculate human segment orientation or joint angles. 85% of these studies (17/20) used a motion analysis system (e.g., Vicon, Qualysis, OptiTrak) as a reference to validate their method. 45% of studies (9/20) used magnetometer data in addition to using the accelerometer and gyroscope signals.

Six studies used the linear Kalman filtering approach to calculate orientation. Two of these studies involved low impact movements [164, 165] and four studies involved no impact movements [167, 172–174]. Two of the six studies calculated RMSE (Root Mean Square Error) for roll and pitch and yaw (including magnetometer), two of the six studies calculated RMSE for roll and pitch (attitude estimation), and two of the six studies calculated average attitude RMSE (from roll and pitch) [See table 3.1]. Of the four studies which calculated RMSE of roll and pitch angles, each study had two distinct levels of testing speeds. The average RMSE of the slow movements in these studies were roll 0.9 degrees (with a range of maximum error of 1.8 degrees and minimum error of 0.1 degrees) and pitch 0.9 degrees (with a range of maximum error of 1.5 degrees and minimum error of 0.05 degrees). The average RMSE of the faster movements in these studies were roll 1.9 degrees (with a range of maximum error of 5.3 degrees and minimum error of 0.1 degrees) and pitch 2.1 degrees (with a range of maximum error of 4.0 degrees and minimum error of 0.05 degrees). In each of these four studies, testing at a faster speed increased the error in roll, pitch and yaw angles. This suggests that the testing speed affects the accuracy of the Linear Kalman Filter. Three of these four studies used testing movements which involved no impacts [172–174], and one study tested their approach on walking which has very low impacts [164], therefore it is not possible to determine the effect of impact using linear Kalman filtering. No studies evaluated their algorithms using running movements, in fact, none of theses studies focused on movements faster than walking. With regard to test durations, none of these studies specifically examined the effect of (long term) measurement duration and no studies carried out testing for periods of greater than 2 minutes.

Only two of the six studies included the use of magnetometer data [172, 174], where only Zhang et al. [172] included a method to compensate for magnetic distortion. While Zhang et al. [172] had the most accurate RMSE for pitch roll and yaw angles, it cannot be concluded that this accuracy is solely related to the inclusion of the magnetometer data because they did not compare the calculations to the accuracy without the magnetometer and they only examined simple rotation movements for a duration of 30 seconds. This is one of the shortest measured testing durations.

Table 3.1: Studies examined using a Linear Kalman Filter.

Author	Test Duration (Seconds)	Sensors	Presence of Impact	Movement Type	Orientation Accuracy		
					Roll (°)	Pitch (°)	Yaw (°)
Tang et al. et al., 2015	30-60	Acc/Gyro/Mag	No	Rotations	1.1	1.5	2.1
				Lunge	1.7	3.6	4.8
Zhang et al., 2012	30	Acc/Gyro/Mag	No	Rotations (slow)	0.1	0.1	0.1
				Rotations (fast)	0.1	0.1	0.1
Mazza et al., 2012	40	Acc/Gyro	Low	Walking (slow)	0.5	0.6	
				Walking (fast)	0.6	0.7	
Lee et al., 2012	30-120	Acc/Gyro	No	Rotations (slow)	1.8	1.4	
				Rotations (fast)	5.3	4.0	
Zhu and Zhou, 2004	25	Acc/Gyro/Mag	No	Rotations	0.2	0.2	1.5
					Average Attitude RMSE(°)		
Luinge et al., 2005,	120	Acc/Gyro	No	Manual Tasks	3.0		
Ligorio and Sabatini,	60-90	Acc/Gyro	Low	Manual Tasks	3.6		
				Walking	1.8		

In practise Kalman filters can be hard to implement, which is reflected in the substantial number of varied Kalman filtering approaches. Though the linear Kalman filter approach can lead to high levels of accuracy, not all estimation problems can be accurately described using a linear model. In these cases, linear modelling of a system is not appropriate and non-linear variants of the Kalman filter should be implemented instead.

3.4.2 Extended Kalman Filter

While the Linear Kalman filter is an optimal estimator for linear systems with Gaussian distributions, but many real-life problems are described by non-linear functions/systems. Other variations of the Kalman filter exist that can be used for these non-linear estimation problems such as the Extended Kalman filter (EKF). The EKF, like the original Kalman Filter, is a two-step estimation problem composed of a prediction step and an update step. This approach implements a solution through linearising a non-linear function around a single point. The EKF uses partial derivatives (Taylor series) to find a linear approximation of the non-linear system around the current estimate (Gaussian mean). The matrix of first-order partial derivatives of the Taylor series expansion is called the Jacobian matrix (which is involved in the linearisation). Full equations describing the EKF approach can be

found in [171].

The EKF appears to be the most widely used variant of the Kalman Filter, and there appears to be many different implementations. The EKF does not lead to an optimal solution when the state transition and measurement model are non-linear (systems of equations), the EKF provides an approximate optimal solution of the estimate. When the state transition and measurement model are linear, the EKF is equivalent to the regular Linear Kalman Filter described in the previous section (Section 3.4).

Nine studies used the EKF to calculate orientation; five studies examined simple rotational movements, two studies used slow movements and two studies involved fast movements (see table 3.2). 45% of studies (5/11) calculated angles for pitch, roll and yaw, while the remaining papers gave a joint average error value. The average RMSE of these studies was 1.1 degrees, 1.0 degrees and 6.5 degrees, for the roll, pitch and yaw angles respectively. Where solely simple rotational movements (no impact) were examined, errors were low (<4 degrees), but these may not be representative of the type of errors achieved when applying the same methods to a faster movement where impact are present in the signal (e.g., walking, running).

Two of the nine studies compared walking to running [175, 176], which provide insight to the combined effect of higher speeds and higher impact. In both studies, error in the average RMSE joint angles were greater for running than walking (by approximately 3 degrees in both studies).

Three studies involved testing during movements involving impacts [12, 175, 176], and two of these studies tested during running specifically [175, 176]. Cooper et al. [175] examined running and showed that for at least 5 minutes of testing (longer than the commonly reported testing durations), the EKF can keep orientation errors reasonably small for running movements. The remaining study which included running, Jakob et al. [176], only tested for duration of 10 seconds. Overall, the duration of experimental testing time varied from 10 seconds to only 5 minutes in these ten studies, and none of these studies directly examined the effect of duration on the resulting accuracy of their proposed approach.

Approximately half of the studies (6/9) made use of magnetometer signal data, but unfortunately none of these studies compared directly the inclusion and exclusion of the magnetometer signal. In addition, these six studies only tested their algorithm on rotation movements (no impact) [6, 177–180] or slow movements (low impact) [12], and so the effect of including the magnetometer was not tested in a fast movement situation or in the presence of high impacts.

Table 3.2: Studies examined using an Extended Kalman Filter.

Author	Test Duration (Seconds)	Sensors	Presence of Impact	Movement Type	Orientation Accuracy		
					Roll (°)	Pitch (°)	Yaw (°)
Sabatini, 2006	2 minutes	Acc/Gyro/Mag	No	Rotations	1.3	1.4	4.13
Labbé et al., 2015	5 minutes	Acc/Gyro/Mag	No	Rotations	0.72	0.78	2.57
Sabatini, 2011	2 minutes	Acc/Gyro/Mag	No	Rotations	0.89	0.97	1.14
Sabatini, 2011	35 seconds	Acc/Gyro/Mag	No	Rotations	0.72	0.83	1.23
					Average Joint Angle °)		
Lin et al., 2012,	-	Acc/Gyro	No	Knee rehabilitation	6.5		
Yun and Bachmann, 2006,	20 seconds	Acc/Gyro/Mag	No	Rotations	9.0		
Cooper et al., 2009,	5 minutes	Acc/Gyro	High	Walking	0.7		
				Running	3.4		
Jacob et al., 2013,	10 seconds	Acc/Gyro	High	Walking	7.0		
				Running	10.2		
				Jumps	7.0		
					Average Attitude Error °		
Luinge and Veltink, 2004,	2 minutes	Acc	No	Manual Tasks	2.0		
Bergamini et al., 2014,	1-3 minutes	Acc/Gyro/Mag	Low	Manual Tasks	3.5		
				Walking	5.5		

There are three main disadvantages of using the EKF. Firstly, most and perhaps most importantly to this thesis, the costly calculation of Jacobian matrices within the algorithm (which are a core part of the EKF) are non-trivial and can be computationally heavy [181]. This suggests that this may not be an appropriate method to consider for real-time human movement tracking. Additionally, the approach of locally linearising a non-linear function can be unstable and introduce errors [181, 182]. Finally, this approach can be difficult to tune and implement which is reflected in the many varied implementations of the EKF.

3.4.3 Unscented Kalman Filter

The approach used by the EKF to tackle non-linear estimation problems is to linearise the non-linear function around a single point, the estimated mean (Gaussian mean). This linearisation can lead to complications which cause filter divergence or poor performance [182], so additional approaches have been investigated. The Unscented Kalman Filter (UKF) is another extension of the original Kalman filter for use in non-linear estimation problems which was proposed by Julier and Uhlmann in 1997 [181]. The UKF, like the original Kalman Filter, is a two-step estimation problem composed of a prediction step and an update step. The UKF implements a solution without the need for linearisation, by defining a sample of points from the original distribution called Sigma points. Sigma points are assigned weights and transformed using a non-linear mapping function to compute a new Gaussian. The mean and variance of the new estimated Gaussian can then be approximated.

The UKF differs from the EKF in that the EKF takes a linear approximation of the non-linear function around one single point, the Gaussian mean, rather than taking a set of a number of sigma
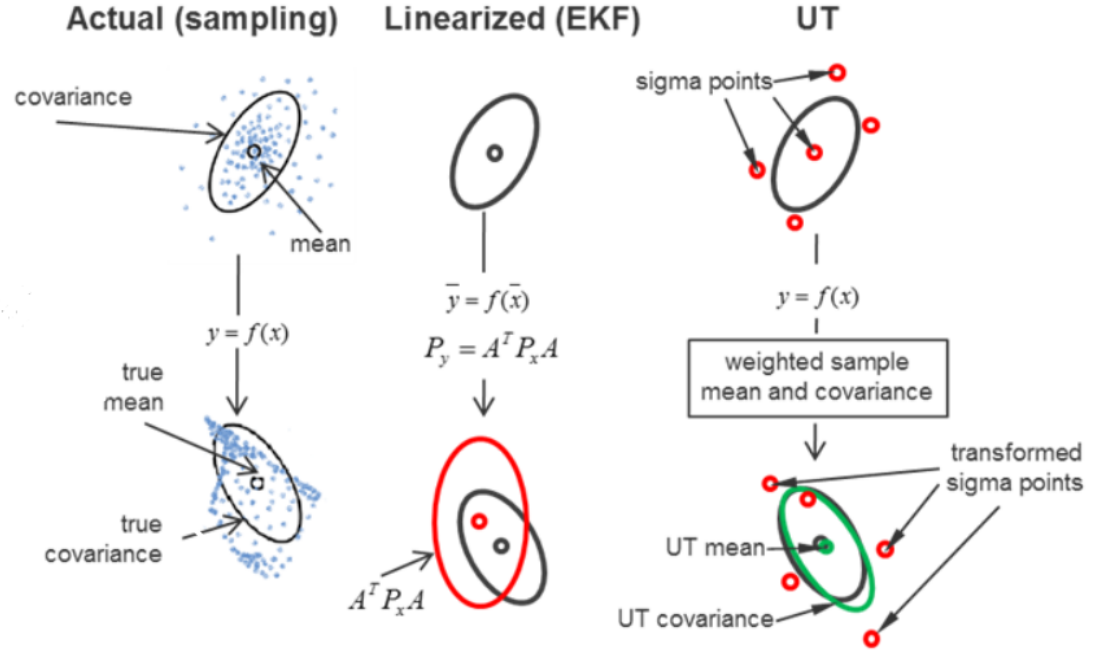
Figure 3.2: Differences between the Extended Kalman Filter and Unscented Kalman Filter approximations.

points (including the mean) which are representative of the whole distribution and passing them through a non-linear function as is performed with the UKF. As show in figure 3.2 this results in a better approximation of the mean and co-variance of the current estimate.

Two studies by El-Gohary and McNames [9, 183] used Unscented Kalman Filtering to estimate orientation, where magnetometer data was excluded. Both studies tested their proposed approach using rotational movements (no impact) with a robotic arm and have given error in terms of RMSE of specific joint angles. El-Gohary and McNames [9, 183] tested their approaches by performing rotation movements at two distinctly different speeds, one fast and one slow. The results from both studies showed that faster movements were associated with higher levels of joint angles estimation error, with average errors less than 6.5 degrees RMSE, compared to average errors less than 8 degrees for faster movements. This potentially highlights the challenge of a decrease in orientation accuracy with higher speed of movements while using the UKF. None of the studies tested the UKF on movements more complex than rotational movements, and no studies evaluated impact related activities. Therefore, it is unclear the appropriateness of this method for running applications.

With regard to the effect of duration, neither studies directly examined this on the resulting accuracy of their proposed approach. However, El-Gohary and McNames (2015) [9], tested their methodology on a fast movement over a 15-minute duration and had average RMSE of 2.7 degrees.

This was the longest testing duration of any study included in this review. However, El-Gohary and McNames (2015) does not allow any conclusion to be drawn on the effect of this method over time for fast and high impact movements.

The UKF is based on the idea that *it is easier to approximate a Gaussian distribution that it is to approximate an arbitrary non-linear function or transformation* [184]. An advantage of the UKF compared to the EKF, is that the need to calculate potentially complicated Jacobian matrices is removed. In addition, the process of local linearisation (which can be unstable) is removed, and instead the UKF uses a deterministic sampling approach to estimate the Gaussian mean and co-variance (with a minimal set of Sigma points).

Table 3.3: Studies examined using an Unscented Kalman Filter.

Author	Test Duration (Seconds)	Sensors	Presence of Impact	Movement Type	Average Joint Angle °	
El-Gohary and McNames, 2015	18 seconds/2 minutes	Acc/Gyro	No	Rotations (slow)	Elbow (flexion/extension)	6.5
				Rotations (fast)	Elbow (flexion/extension)	8.0
l-Gohary and McNames, 2015	15 minutes	Acc/Gyro	No	Rotations (slow)	Elbow (flexion/extension)	0.9
					Shoulder (flexion/extension)	0.8
					Wrist (flexion/extension)	1.1
				Rotations (fast)	Elbow (flexion/extension)	2.8
					Shoulder (flexion/extension)	2.5
					Wrist (flexion/extension)	1.8
					Average Error (°)	
LaViola, 2003	20 seconds	Acc/Gyro	No	Rotations	Head	0.32
					Hand	0.32
					Orientation Accuracy	
					Roll (°)	Pitch (°)	Yaw (°)
Momi and Ferrigno, 2015	5 seconds	Acc/Gyro	No	Rotations	0.29	0.32	0.43

3.5 Complementary Filter

Complementary filtering is a sensor fusion approach which involves filtering signals based on certain complementary characteristics of their frequency domain and fusing the data from these filtered signals together to produce stronger combined information [185], e.g., a better estimate of orientation. Typically, this involves filtering the accelerometer and magnetometer data using a low pass filter, to remove the high frequency noise from these signals [186]. Similarly, gyroscope data is filtered with a high pass filter, to remove low frequency noise due to biases or drift [186].

There are many ways to implement a complementary filter, the concept is the same, but the filter design may vary. Certain parameters associated with the algorithm can be tuned or adjusted for the target application. For example, the Madgwick filter [14] is a form of complementary filter, which contains one tuning parameter which dictates the convergence of the algorithm. Many studies set the value of this parameter, to the value employed by the original implementation, but the appropriateness

of the value chosen may depend on the target application and measurement error of the gyroscope used. Additionally, the Madgwick Filter algorithm contains an optimisation process called Stochastic Gradient Descent. It is possible Stochastic Gradient Descent may itself be improved upon with a different optimisation process.

Fourteen studies have been included in this review for the purpose of providing a scope of the area of Complementary Filtering to calculate orientation (see table 3.4). Approximately 64% of these studies (9/14) used a motion analysis system (e.g., Vicon, Qualysis, OptiTrak) as a reference to validate their method. 72% of studies (10/14) used magnetometer data as well as the accelerometer and gyroscope signals (MARG approach), 14% of studies (2/14) accelerometer and gyroscope signals only (IMU approach), and 14% of studies (2/14) implemented both a MARG and IMU approach separately.

Of the fourteen papers, eleven studies tested the orientation accuracy of no-impact movements (e.g., simple rotational movements [14, 186–194] and cycling [195]), one study tested the orientation accuracy of low impact movements (e.g., walking [196]), and two studies tested the orientation accuracy of high impact movements (e.g., running [8, 197]).

Of the fourteen papers, three studies used fast movement speed (during running [8, 197] and cycling [195]) and one paper used moderate movement speed (during walking [196]). Of the three studies evaluating fast movements, Cockcroft et al. [195] evaluated hip joint angles during cycling and reported mean absolute errors of less than 3 degrees at low, medium, and high pedalling speeds. When comparing the joint angle errors associated with slow, medium, and fast pedalling, there was an increase in error as pedalling speed increased. Unlike other studies which examined walking or running, where speed and impact levels are related [196], cycling does not involve impact so the increase in error reported here is due to speed alone. Shull et al. [8] estimated trunk orientation during walking and running, and also reported an increase in error associated with the faster movements. Walking with normal trunk movement was tested at two walking speeds (similar levels of impact but faster speed), and the faster walking speed resulted in higher levels of inaccuracy compared to the slower walking speed, with similar results reported in Huang et al. [196]. Running with normal trunk movement was also tested in Shull et al. [8], which had higher levels of inaccuracy compared to both tested walking speeds. Kim et al. [197] valuated their complementary filer for one level of walking speed, and one level of running speed, and their results were in agreement with Shull [8], where the orientation accuracy associated with running had greater error than the orientation accuracy associated with walking. Shull

et al. [8] and Kim et al. [197] were the only studies reviewed to evaluate a high impact movement, running.

Table 3.4: Studies examined using an Complementary Filter.

Author	Test Duration (Seconds)	Sensors	Presence of Impact	Movement Type	Algorithm		Mean Joint Angle °	
Kim et al., 2015,	-	Acc/Gyro/Mag	High	Walking and Running	Complementary filter	Walking	2.0	
						Running	2.8	
Karunarathne et al., 2014	Acc/Gyro	200 seconds	No	Rotations	Complementary filter		8.7	
							Mean Absolute Error °	
Cockcroft and Scheffer, 2014	Acc/Gyro/Mag	15 minutes	No	Cycling	Pendulum Filter	Slow	2.8	
						Medium	2.7	
						Fast	2.5	
					Xsens	Slow	2.1	
						Medium	2.6	
						Fast	2.6	
					Passive Filter	Slow	4.3	
						Medium	9.4	
						Fast	19.8	
							Mean Segment Angle Error °	
Sarbishei, 2016	Acc/Gyro	100 seconds	No	Rotations	Madgwick IMU		0.3	
					OMID IMU		0.2	
Shull et al., 2016	Acc/Gyro/Mag	2 minutes	High	Walking and Running	Magneto-Gyro	Walking 1.2m/s	1.1	
						Walking 1.6m/s	1.3	
						Running 2.4m/s	1.8	
					Madgwick IMU	Walking 1.2m/s	1.4	
						Walking 1.6m/s	1.4	
						Running 2.4m/s	3.3	
Huang et al., 2016	Acc/Gyro/Mag	2 minutes	Low	Walking	Magneto-IMU	Walking 1.0m/s	1.9	
						Walking 1.2m/s	2.1	
						Walking 1.4m/s	2.5	

Author	Test Duration (Seconds)	Sensors	Presence of Impact	Movement Type	Algorithm	Orientation Accuracy		
						Roll (°)	Pitch (°)	Yaw (°)
Madgwick et al., 2011	Acc/Gyro/Mag	60 seconds	No	Rotations	IMU Static	0.5	0.6	-
					MARG static	0.5	0.6	1.1
					IMU dynamic	0.7	0.6	-
					MARG dynamic	0.7	0.6	1.1
Kubelka and Reinstein, 2012	Acc/Gyro	15 minutes	No	Rotations	Complementary Filter	2.8	4.1	
Sarbishei, 2016	Acc/Gyro	100 seconds	No	Rotations	Madgwick MARG	0.4	1.9	1.8
					OMID MARG	0.2	1.0	0.9
Fourati, 2012	Acc/Gyro/Mag	60-80s	No	Rotations	Complementary filter	3.0	3.0	4.0
Fan et al., 2017	Acc/Gyro/Mag	60-80 seconds	No	Rotations	Complementary Filter	1.7	1.8	1.9
					Madgwick IMU	3.6	2.8	3.0
					Madgwick MARG	4.5	3.3	3.2
Tian and Tan, 2012	Acc/Gyro/Mag	17 minutes	No	Rotations	Proposed Filter	0.2	0.3	0.5
					Madgwick MARG	0.6	0.5	1.1
Tian and Tan, 2012	Acc/Gyro/Mag	-	No	Rotations	Proposed Filter	0.7	0.6	0.8
					Madgwick MARG	2.8	2.7	10.1
Rasteiro et al., 2015	Acc/Gyro/Mag	30 seconds	No	Rotations	AGMDF	6.1	2.7	6.4
					AGMHF	6.2	2.7	6.4
					OMDF	6.1	3.6	6.7
					AMHF	7.5	3.7	6.4
Rasteiro et al., 2015	Acc/Gyro/Mag	30-60 seconds	No	Rotations	AGMDF	5.0	3.6	7.8
					AGMHF	5.5	3.8	7.8
					OMDF	7.9	5.0	9.8
					AMHF	11.8	7.7	13.8

With regard to test duration none of these studies directly examined the effect of (long term) duration on the resulting accuracy of their proposed approach. In general, the studies were tested for durations up to only 3 minutes, with the exception of three studies which tested simple rotations [187, 192] and bicycle pedalling [195] for approximately 15 minutes, with average RMSE <4 degrees in both studies. Of the studies which examined high impact movements, Shull et al. [8] captured data for less than 2 minutes, which is not sufficiently long enough to determine the long-term accuracy of their approach, and Kim et al. [197] did not include explicit information on the duration of data capture.

It appears some of the most accurate implementations of the complementary filter are modified versions of the Madgwick filter [188, 192]. However, the approach in Tian and Tan [192] relied on the

magnetometer signal as part of their algorithm design, and Sarbishei's filter [188] has been designed for commercial use and embedded in the Neblina motion sensor [198].

The Complementary filter has been commonly adopted as an alternative to other orientation estimation methods due to it being a computationally light solution to the orientation estimation problem, being relatively easy to implement [162, 197, 199]. For these reasons the Complementary filter is one of the most popular orientation estimation approaches.

3.6 Deterministic Methods

Deterministic methods [200] (e.g., TRIAD, QUEST [201], and FQA [202]) are additional approaches to estimate orientation, which were proposed as solutions to Wahba's optimisation problem [203]. Detailed descriptions of these approaches can be found in can be found in [204]. These methods provide an absolute estimate of orientation using vector observations from two or more measurement vectors, e.g., accelerometer and magnetometer. These methods only perform orientation estimation well for static and slow movements. This is due to the fact that only the accelerometer and magnetometer are used for the orientation tracking, so any dynamic movements will cause high frequency noise in the accelerometer measurement vectors and any local magnetic field disturbances will distort the magnetometer measurement. The presence of either of these conditions will deteriorate the ability of these deterministic approaches to accurately track orientation [6, 199]. Deterministic methods have been shown to be outperformed by both Stochastic filtering and Complementary filtering approaches which typically involve the same deterministic estimation but with additional steps to compensate its disadvantages [191, 192, 204]. As a result, these deterministic approaches are not implemented often as a sole means to estimate orientation [205], but they have been used jointly with Stochastic or Complementary approaches mentioned above [199, 202, 204, 206]. For these reasons these methods were not included in the detailed comparative section (Section 3.6) of this review.

3.7 Direct Comparison of Kalman Filtering and Complementary Filtering

Based on evaluation of the approaches described above and current theory, it is clear that the Kalman filtering approaches and Complementary filtering are the commonly used in human movement analysis. However, in order to determine which is most appropriate for the current thesis with running analysis

as a focus, it is necessary to examine studies which have directly compared the two approaches.

This section highlights studies which have directly compared Kalman Filtering (Linear Kalman Filtering and Extended Kalman Filtering) and Complementary Filtering to each other. It is recognised that both the Extended Kalman Filter (EKF) and the Unscented Kalman Filter (UKF) are both used for non-linear estimation with the Kalman Filter. However, Rhudy et al. [207] concluded after an extensive review of the two approaches, that while studies have reported mixed conclusions on the comparison of the EKF to the UKF, with similar levels of accuracy reported for human based estimation [208], the EKF is recommended for real-time applications due to the lower amount of computation involved compared to the UKF [207, 208]. Due to this conclusion, the EKF is the non-linear Kalman filter included in this review. Given that this PhD relies on an algorithm which does not use magnetometer signal data (see Section 1), this review will not include the findings from studies which require magnetometer data. In comparing Complementary filtering and Kalman Filtering for running, it is worth evaluating the studies reviewed in terms of orientation accuracy, the speed of the captured movement, the level of impact of the captured movements and the computational cost of the approach.

Eight studies have directly compared 6DOF implementations of Kalman Filtering and Complementary Filtering with regard to orientation accuracy. Of these studies, 4 have compared implementations of a Linear Kalman filter and a complementary filter [8, 14, 163, 209], 3 studies have compared an EKF to a Complementary filter [15, 210, 211], and one study compared both Linear and Extended Kalman filtering to a complementary filter, Shan and Wu [16]. The Madgwick filter [14] and Mahony filter [212] are the most common implementation of Complementary filters, in half of the 8 studies either one or both of these algorithms were compared to a Kalman filter.

One of the 8 studies, Shull et al., identified the Complementary filtering approach to be more accurate than a Linear Kalman Filter [8]. Two studies identified an Extended Kalman filtering approach to be more accurate than the Complementary filter used (Valencia and Vargas [211] and Sihite and Bewley [15]). The remaining 5 studies [14, 16, 163, 209, 210] reported results that were close in accuracy between the two filtering approaches with Kalman filtering being slightly more accurate in 4 of the five studies [16, 163, 209, 210]. The mean $\pm$ standard deviation of the difference of measured dynamic accuracy was 0.4 ± 0.3, 0.4 ± 0.5 and 0.5 ± 0.4 for roll, pitch and yaw, respectively. The full list of studies review can be found in table 3.5.

Table 3.5: Comparative view of studies included Direct Comparison of Kalman Filtering and Complementary Filtering.

Author	Test Duration (Seconds)	Presence of Impact	Movement Type	Algorithm	Mean Joint Angle °		
Valencia and Vargas, 2020	15 seconds	No	Rotations	Complementary Filter	5.7		
				Extended Kalman Filter	3.3		
					Mean Attitude RMSE °		
Sihite and Bewley, 2018	35 seconds	No	Static	Complementary Filter	1.3		
				Extended Kalman Filter	0.8		
			Rotations	Complementary Filter	0.8		
				Extended Kalman Filter	1.4		
					Mean Segment Angle Error °		
Shull et al., 2016	2 minutes	High	Walking and Running	Madgwick IMU	Walking 1.2m/s	1.4	
					Walking 1.6m/s	1.4	
					Running 2.4m/s	3.3	
					Orientation Accuracy		
					Roll (°)	Pitch (°)	Yaw (°)
Harrison and Vaidyanathan, 2011	60 seconds	No	Rotations	IMU static	0.5	0.6	-
				Kalman Filter static	0.8	0.8	1.2
				IMU Filter dynamic	0.7	0.6	-
				Kalman Filter dynamic	0.9	0.8	1.3
Ohberg and Grip, 2013	60 seconds	No	Rotations	Complementary Filter	0.8	0.6	2.0
				Kalman Filter	0.1	0.1	1.7
		Yes	Walking	Complementary Filter	1.3	0.3	0.1
				Kalman Filter	0.3	0.2	0.3
Giurato and Lovera, 2016	60 seconds	No	Drone flight	Complementary Filter	1.2	1.1	1.5
				Extended Kalman Filter	0.8	0.9	0.7
Shan and Wu, 2017	35 seconds	No	Static	Complementary Filter	0.3	0.2	
				Linear Kalman Filter	0.1	0.2	
				Extended Kalman Filter	0.2	0.2	
				Mahony Filter	0.2	0.2	
			Dynamic	Complementary Filter	2.4	1.9	
				Linear Kalman Filter	2.3	1.9	
				Extended Kalman Filter	8.0	11.0	
				Mahony Filter	6.8	6.8	
Htun et al., 2018	50 seconds	No	Rotations	Complementary Filter	-	4.8	
				Kalman Filter	-	3.3	

3.7.1 Linear Kalman Filter vs Complementary Filter

Five studies have examined a comparison between the Linear Kalman filter and a Complementary filter. Three of the five studies reported that the Linear Kalman filter was more accurate when compared to the Complementary filter [16, 163, 209]. The average RMSE differences of roll, pitch and yaw between the two approaches in these 3 studies were 0.6 ± 0.6 (mean $\pm$ SD), 0.6 ± 0.8 and 0.2 ± 0.3 respectively. Three of the five studies implemented the Madgwick filter as their Complementary filter [8, 14, 16], of which the two remaining studies reported that the Complementary filter was more accurate compared to Linear Kalman filter [8, 14].

Of the five papers identified, three studies tested the algorithms on a movement other than simple rotations [8, 16, 209]. The approach in Shan and Wu [16] was tested by quickly moving the inertial sensor horizontally across a table to introduce high accelerations. The results in their study indicated that the Complementary filter and the Kalman Filter were comparable for the dynamic conditions with an RMSE difference of 0.2 degrees on the roll axis, and 0.1 degrees on the pitch axis.

Ohberg et al. [209] and Shull et al. [8] both tested their approaches on two distinct types of movement, with Ohberg [209] comparing simple squatting (segment rotations) to walking, and Shull

[8] comparing walking to running. The algorithm accuracy in Ohberg et al. [209] stays relatively consistent for both the Kalman filter and Complementary approach between the two movements (walking and squatting). It could be that the acceleration due to the movements are not strong enough to cause any additional error in movement tracking with (low speed and low impacts in walking and no impacts in squatting), considering the accuracy of the two approaches are so close. Whereas in Shull et al. [8], the increase in orientation error is directly related to an increase in gait speed, and distinctions can be made between the approaches and their ability to handle the increased impacts and speed; with the Madgwick Complementary filter more capable at handling the noisier sensor data, in this example. Shull et al. [8] appears to be the only study which compares a 6DOF implementation of a complementary filter with other orientation estimation algorithms (i.e., Kalman filtering) during running specifically, with average segment angle errors of 3.3 degrees. It is accepted that the exclusion of the magnetometer is detrimental in terms of handling signal drift on the yaw axis, but Shull et al. show that smaller levels of inaccuracy may be possible using 6DOF in comparison to 9DOF, as the 9DOF Kalman implementation in Shull et al. [8] using (accelerometer, gyroscope and magnetometer) performed worse than the 6DOF complementary filter. Although it should be noted that the sensor in Shull et al. [8] is placed on the spine and would experience higher loads of acceleration when placed on the lower body. It is interesting to note that the complementary filter used in [8] (Madgwick filter) was more accurate than the Kalman filter used which included a magnetometer. Additionally, the Madgwick filter was less accurate than another complementary filter also examined, which consisted of a magnetometer and a gyroscope only. This shows the level of inaccuracy was linked to the inclusion of the accelerometer. But given the focus of this research is to exclude the issues related to the magnetometer, an accelerometer must be used to compensate for the errors in gyroscope orientation estimation.

3.7.2 Extended Kalman Filter vs Complementary Filter

Four studies examined a comparison between Extended Kalman filtering and Complementary filtering, Giurato et al. [210], Valencia et al. [211], Sihite et al. [15], and Shan et al. [16] (where Shan et al. [16] compared both Kalman filtering approaches). Three of the four studies suggested that the EKF was more accurate than the Complementary filtering method [15, 210, 211]. Comparing the difference in accuracy between the two approaches for these studies, Valencia and Vargas [211] had 2.4 degrees difference in joint angle, Sihite and Bewley [15] had 7.8 degrees difference in average dynamic roll and

pitch, and giurato and Lovera [210] had 0.4, 0.1, and 0.8 degrees difference in RMSE for dynamic roll, pitch and yaw, respectively. The difference in accuracy between EKF and Complementary filter are larger than in the comparison with Linear Kalman filtering. This is also reflected in the remaining study where the Complementary filter was more accurate, with differences in RMSE for dynamic roll and pitch were 5.6 and 9.1 degrees respectively.

In terms of the movement captured, Shan and Wu [16] was already discussed in the previous section with the Linear Kalman filters. In the remaining three studies [15, 210, 211], no algorithms were tested on movements more dynamic than simple segment rotations, including Giurato and Lovera [210] which was tested by hovering a drone where specifically no quick manoeuvres were performed. Additionally, none of these studies involved the presence of any impacts. This makes it difficult to assess the orientation accuracy this approach for a human movement like walking or running. Valencia and Vargas [211] performed on a human movement specifically calculating knee angle.

3.7.3 Additional Factors to Consider

When considering all studies reviewed in this section, the average length of data capture was approximately 60 seconds, within which most of these studies had consistently low errors in orientation accuracy (see Table 3.5). However, as none of these studies tested their captured movements for longer than 2 minutes, it is hard to estimate the performance which would be found over longer durations, knowing inaccuracy is likely to increase with time and movement speed. For true evaluation of how these algorithms would perform for long-term human movement analysis such as running, the approaches need to be tested on human movements for longer periods of time e.g., >30 minutes and for movements with higher levels of impacts.

For real-time applications the computational time of the method chosen is important, as mentioned in a number of 6DOF [16, 163, 209], and 9DOF orientation estimation studies [197, 213–215]). These studies highlight the large overhead for the (Linear and Extended) Kalman filters, with the complementary filter in Young [199] performing up to nine times faster than the EKF used in their study. Young suggests that the poor performance of their Kalman filter is due to the process model used being a poor approximation of the dynamics of leg walking. This highlights the high dependence of the Kalman filtering approach on an accurate process model, which is difficult to design for human movement due to the various dynamics of different body segments [199].

An important aspect of 6DOF orientation estimation is the presence of disturbances from external

acceleration, when the orientation of a moving person/object is considered. During 6DOF orientation estimation the gravity vector from the accelerometer is used as the sole reference for estimating attitude (pitch and roll angles). During slow movements, this vector can be clearly read, as the gravity represents the majority of the motion measured via the accelerometer. However, during faster movements, the motion measured by the accelerometer is a combination of both gravity and the faster body accelerations (accelerations of the moving person/object) which distorts its accuracy as a reference for orientation. It is important that this error is reduced as much as possible. A recent study by Parikh et al. [216] comparing the accuracy of the Linear Kalman filter, Madgwick filter, the Mahony filter under external accelerations using a quadrotor, found that the Madgwick filter performed the best to mitigate the errors due to external accelerations.

3.8 Conclusion

In light of the proposed PhD study to examine running over an extended time period (e.g. 30 minutes) using 6DOF only, this review highlighted a selection of studies which compared two possible approaches: Complementary and Kalman filtering. Examples of representative studies can be found in tables 3.5. On evaluation of the orientation accuracy of these approaches within the reviewed studies, the majority of studies (5/8) had relatively similar levels of accuracy. Given that the aim of this research is to evaluate running in for a real-time application, Complementary filtering is suggested to be more appropriate. The increased accuracy of the Kalman filtering approach for the additional computational overhead does not seem advantageous in this case. In addition, it has been shown that the Madgwick filter can be more effective than other approaches at limiting the effects of external accelerations in certain test conditions [216].

Based on this review, it can be seen that there is a need for further research to examine the level of accuracy of calculating human segment orientation during for an extended period of time. The advantage of 9DOF orientation estimation is known, however 9DOF estimation may not always be suitable for the target application and a 6DOF approach should be investigated. Only one study was identified which calculated 6DOF orientation for running where accuracy was compared to other approaches [8].

The most appropriate method to use in this thesis appears to be the Complementary Filtering approach, as it is comparable in accuracy with the EKF [214], but easier to implement and less computationally heavy for real-time applications [162, 197, 199]. However, it is unclear the level of

accuracy of the Complementary Filter with regard to high impact, fast, running movements for extended periods of time.

Chapter 4

Related Work: Biofeedback

4.1 Use of Biofeedback in Running

In light of the large number of injuries associated with running [5], and the relationship between high loading and risk of injury [27], there has been a growing interest in altering running technique to reduce loading on the body [20, 217]. Running gait re-training may be targeted at healthy individuals who wish to avoid injury or individuals who wish to avoid re-injury. A recent focus has centred around the use of technologies to provide real-time biofeedback. Biofeedback is an approach by which instantaneous *augmented feedback* (verbal, visual, auditory or haptic sources) is provided in order to give additional quantifiable information. In the case of running, this augmented feedback generally relates to a runner's loading or technique. Given the focus of this thesis is to examine RRIs associated with high loading and inappropriate technique, the following sections outline: an introduction to biofeedback, an examination of the use of loading-based (vGRFs and acceleration) biofeedback and technique-based biofeedback in running, and finally a discussion on the difficulties that exist in existing biofeedback research.

4.1.1 Introduction to Biofeedback

The two types of augmented feedback provided to a runner can be broadly characterised as either knowledge of results (KR) or knowledge of performance (KP) based feedback. KR refers to feedback associated with outcome-based measures (i.e., impact accelerations, running speed) whereas KP refers to feedback associated with technique (i.e., thigh flexion angle, knee flexion angle). There is limited evidence to conclude whether KR or KP is more effective for the retention of new running re-training techniques; both KR and KP have separately been proven to be effective forms of biofeedback to reduce loading during running (see tables 4.2 and 4.3). However, only one study appears to have directly compared whether KR or KP based feedback is more effective at providing technique retraining during

running [24]. Although Baggaley et al. [24] did not explicitly focus on the concepts of KR and KP, the comparison of the effectiveness of different feedback parameters provided could be categorised as KR (VALR [24]) versus KP (foot strike pattern, step length [24]). Baggaley et al. [24] found that measures of both KR and KP based biofeedback lead to significant reductions in loading [measured as Vertical Average Loading Rate (VALR)], however, the provision of KP based feedback (specifically modified foot-strike patterns) lead to greater reductions in loading than KR based VALR biofeedback (47% reduction versus 31%, respectively).

Within the broader research area of general motor learning, the relative effectiveness of KR and KP for improving motor learning is unclear due to limited and conflicting evidence [218, 219]. KR and KP have both been shown to be an effective source of biofeedback in non-running studies [218–220]. Historically, the majority of motor learning research investigating the effectiveness of augmented biofeedback has focused on KR evaluating simple movements performed in laboratory environment [219, 221]. However, recent comparisons have highlighted the effectiveness of KP over KR in for certain applications [218, 220, 222].

In relation to the area of running technique retraining, existing studies cover a range of biofeedback modes (visual, auditory, haptic, verbal), with visual feedback being by far the most common form [25, 223]. Often when the main mode of biofeedback is non-verbal (visual, auditory, haptic), there will be verbal instructions provided initially to instruct a runner how to achieve lower levels of loading [224–226]. Different forms of feedback can result in different technique changes [227] and different modes of feedback may be more suited to certain applications and/or environments. For example, although visual modes of biofeedback have performed very well [20, 23, 228], visual feedback for running technique re-training is not particularly useful in outdoor running environments where runners need to pay attention to external distractions such as pedestrians and traffic. In addition, a comparison of visual and auditory modes of biofeedback to reduce impact loads during running found that runners preferred the auditory feedback over visual feedback as it allowed them to focus on their limb movements [229].

4.1.2 Use of Loading-based Biofeedback in Running

Vertical Ground Reaction Forces

Providing biofeedback on loading-based variables in running has been extensively investigated; with vGRFs and peak tibial acceleration the two main loading biofeedback variables employed. The captured vGRFs represent the whole body loading experienced during running [Vertical Impact Peak

(VIP), Vertical Average Loading Rate (VALR) and Vertical Instantaneous Loading Rate (VILR)] (see vGRF curve in figure 2.1). Six studies have been included for review to provide a scope of the use of vGRF loading-based biofeedback in running (Table 4.1). The majority of these studies have provided visual based biofeedback with the remaining studies focusing on auditory biofeedback. Each of these studies showed the ability to significantly reduce vGRF loading variables using directly related biofeedback based on vGRF impact peak [20], VALR [24, 230], and impact sound intensity [228, 231, 232]. The results show that providing biofeedback based on vGRF can lead to reductions in VIP, VALR, and VILR of 4-28%, 12-58% and 15-34%, respectively.

Strong evidence of the benefits using of real-time vGRF biofeedback to reduce loading during running is evident in a recent large scale (n=320) prospective study by Chan et al. [20]. This prospective study examined loading rates (VIP, VALR, VILR) before and after a 2-week treadmill-based running technique re-training program. Biofeedback based on vGRFs (impact peak) were visually presented to the runner during several sessions of running re-training. It was found that there was a significant reduction in all measured loading variables after running re-training, and furthermore, when RRI incidence was tracked over a 1-year period, the group that undertook running re-training had a 62% lower risk of injury compared to the control group. This provides strong justification for the provision of biofeedback to reduce loading during running.

Table 4.1: Selection of studies examining the use of vGRF based real-time biofeedback during running.

Author	Year	Mode of Feedback	No. of Subjects	Training Environment	Feedback Parameter	Outcome Measure	Loading Results
Chan et al., [20]	2017	Visual	166, cont: 154	Treadmill	VIP	VALR VILR	Approx. -21%* across speeds Approx. -15%* across speeds
Baggaley et al., [24]	2016	Visual	32	Treadmill	VALR	VALR	-31%*
Garofini et al., [230]	2020	Visual	20	Treadmill	$VALR_{Init}$ $VALR_{PostKP}$	VALR VALR	-40%* -58%*
Tate & Milner [228]	2017	Visual	14	Treadmill	Impact Sound Intensity	VIP (OG) VALR (OG) VILR (OG)	-28%* -36%* -34%*
Rothstein [231]	2020	Auditory	15	Treadmill	Impact Sound Intensity	VIP (OG) VALR (OG) VILR (OG)	-4%* -12% -15%*
Phan et al., [232]	2016	Auditory	26	Overground	Impact Sound Intensity	VIP VALR VILR	Reduced* Reduced* Reduced*

: Significant finding, Cont: Number of Controls, OG: Overground running, Reduced - Significantly reduced but no values provided, $VALR_{Init}$: Initial VALR biofeedback, $VALR_{PostKP}$: VALR biofeedback directly after technique biofeedback.

Impact Accelerations

One disadvantage of measuring loading via vGRFs (at a whole-body level) is that this does not represent the magnitudes of loading that are present at individual body segments, where injuries may occur. Accelerometers have been used as a proxy measure for the magnitude of local segmental loading

during level running [27, 79] and can be used outside of a laboratory environment. Accelerometers are deemed appropriate because acceleration is proportionate to force (F = ma). A large number of studies have explored the use of impact accelerations (see table 4.2), and specifically peak tibial acceleration (PTA), as a source of biofeedback. Sixteen studies have been included for review to provide a scope of the use of acceleration based biofeedback in running. The majority of these studies have provided visual based biofeedback with the remaining focusing on auditory, verbal, and haptic biofeedback. Each study was able to achieve reductions in the measured loading variables through the use of either PTA, Peak Foot Acceleration (PFA), Resultant Tibia Acceleration (RTA), or segment acceleration (tibia, sacrum and head) based biofeedback. The results show that providing biofeedback based on accelerations can lead to reductions in PTA, VIP, VALR, and VILR of 6-50%, 16-39%, 15-39% and 6-60%, respectively. Each study was capable of a minimum reduction of 10% across loading variables. It can be seen that there is overall a similar percentage reduction of loading achieved using visual and auditory modes of PTA biofeedback, with neither mode out-performing the other. With respect to the studies which examined overground running, there was a 11-44% reduction in PTA, which is in line with treadmill-based methods (6%-32%).

Table 4.2: Selection of studies examining the use of impact acceleration based real-time biofeedback during running.

Author	Year	Mode of Feedback	No. of Subjects	Training Environment	Feedback Parameter	Outcome Measure	Loading Results	Acceleration_pre (g) Mean ± Std Dev.	Acceleration_After (g) Mean ± Std Dev.
Van et al. [233]	2018	Auditory	5	Overground	PTA	PTA	-26%*	10.4 ± 1.5	7.7 ± 0.8
Van et al. [234]	2021	Auditory	10	Overground	PTA	PTA	-27%*	11.1 ± 1.8	8.2 ± 1.8
Van et al. [235]	2021	Auditory	10, Cont: 10	Overground	PTA	PTA	-26%*	10.9 ± 2.8	8.1 ± 3.9
Creaby and Smith [236]	2016	Verbal vs. Visual	22	Treadmill	Clinician vs PTA	PTA	-(19-29%)*	approx. 5.5 ± 3.5	4.1 ± 2.8
Gray et al. [237]	2012	Verbal vs. Visual	8	Treadmill	PTA	PTA	Visual -16%* Verbal -22%*	3.9 ± 1.5[b] 4.4 ± 1.6[b]	3.3 ± 1.2[b] 3.4 ± 1.6[b]
Zhang et al. [238]	2019	Visual	15	Treadmill	PTA	PTA	-29%* (Treadmill) -12%* (Overground)	approx. 12.0 ± 3.0 approx. 10.5 ± 2.0	approx. 7.5 ± 1.5 approx. 9.7 ± 1.5
Crowell et al. [226]	2010	Visual	5	Treadmill	PTA	VIP VALR VILR PTA	-(16%-39%)* -(15-39%)* +6 to -60%* -(6-30%)*	- - - -	- - - -
Crowell and Davis [23]	2011	Visual	10	Treadmill	PTA	PTA VIP VALR VILR	-44%* (Overground) -20%* (Overground) -27%* (Overground) -30%* (Overground)	approx. 8.0 ± 2.5 - - -	approx. 4.2 ± 1.7 - - -
Clansey et al. [229]	2014	Visual, Auditory	3, Cont: 10	Treadmill	PTA	PTA VALR VILR	-31%* -18%)* -19%*	10.7 ± 1.9 - -	7.4 ± 1.2 - -
Wood and Kipp [239]	2014	Auditory	9	Treadmill	PTA	PTA	-10%*	5.9 ± 0.7	5.4 ± 0.7
Bowser et al. [21]	2018	Visual	19	Treadmill	PTA	PTA VIP VALR VILR	-32% -21%* -25%* -37%*	- - - -	- - - -
Sheerin et al. [240]	2020	18	Haptic	Treadmill Overground	Resultant Tibial Acceleration Resultant Tibial Acceleration	VIP	-50%* -28%*	14.8 (12.5, 16.6)[c] 19.4 (17.3, 20.6)[c]	7.4 (6.1, 9.2)[c] 14.0 (11.8, 17.7)[c]
O'Cathain et al. [26]	2020	Verbal	20	Treadmill	tibia, sacrum, head accelerations	VIP Sacrum Head Tibia	-17% -41%* -28%* (no difference)	- - - -	- - - -
Ching et al. [241]	2017	Auditory	16	Treadmill	PTA	PTA PFA VALR VILR	-21% -34%* -25%* -22%*	5.5 ± 1.6 19.0 ± 7.7 - -	4.3 ± 1.5 12.6 ± 6.9 - -
Cheung et al. [242]	2017	Visual	16	Treadmill	PFA	PFA VALR VILR	-32%* -19%* -13%*	11.9 ± 4.9 - -	8.07 ± 2.7 - -
Morris et al. [36]	2019	Auditory, Verbal	128	Treadmill	PTA	no. of foot-strike runners	80% of subjects transitions from RFS to NRFS	-	-

*: Significant finding, Cont: Number of Controls, Std. Dev.: Standard Deviation, PTA: Peak Tibial Accelerations, PFA: Peak Foot Accelerations, RFS: Rear foot-strike pattern, NRFS: Non-rear foot-strike pattern, [b]: Median ± Standard deviation, [c]: Median (Interquartile range),

The tibia can experience high levels of impact loading during running and is often a site of lower limb injury (e.g., tibial stress fractures [27]). Therefore, monitoring loading levels at the tibia is important, however more research is required with respect to running re-training and the use of acceleration-based feedback to measure the dispersion of impact loading further up the body, i.e., at the sacrum, an additional site of injury (lower back strain [243]). Segments such as the pelvis and the trunk have been more commonly examined in running re-training with respect to kinematics, but not loading variables [225, 244–246]. O'Cathain et al. [26] assessed load reduction at the tibia, sacrum and head. They found that although there were no significant reductions at the tibia, the verbal instruction to run softer provided sufficient guidance to reduce loading at the head (28%) and sacrum (41%), showing the effect of running retraining further up the body. Interestingly, measuring the effect of running re-training on loading rates in the vGRF (VALR or VILR) is a common feature of most running re-training studies (see tables 4.1 and 4.2) due to its potential relationship with injury [20]. However, no studies appear to have measured the effect of biofeedback on the rate of loading at a segmental level, as assessed through the rate of acceleration. This may be an important as loading rates have previously been found to be more related to injury at a whole-body level, than impact peaks [1].

4.1.3 Use of Technique-based Biofeedback in Running

The general aim of targeting changes in running technique through running re-training is for runners to adopt a more compliant movement pattern, which is more cushioned from impact loads at foot-strike. Verbal instructions are typically provided in the initial stages of running re-training (alongside the main form of biofeedback [visual, auditory]) as a guide to give runners some direction as to how they can improve their movement technique. These instructions may be as simple as 'run softer' [226] or as specific as ' shorten stride length' [224] or 'run with your kneecaps pointing straight ahead' [246]. A number of running studies have explored the use of technique information as a source of biofeedback. Modification of foot-strike patterns and step rate are the most investigated forms of providing biofeedback on movement technique directly. Twelve studies have been included for review to provide a scope of the use of kinematic/technique-based biofeedback in running. All studies were laboratory-based on a treadmill, with the majority providing visual based biofeedback, followed by auditory, or a mix of biofeedback modes. The results show that providing biofeedback based on movement technique can lead to reductions in PTA (11-13%), VIP (5%), VALR (15-58%), VILR

(14-32%), hip adduction angle (23-29%), and hip internal rotation (17-23%) and increases in step rate (6-9%). One study that compared multiple biofeedback measures (KR: VALR vs KP: foot-strike pattern, step rate), found that a knowledge of performance-based feedback (adopting a non-rearfoot strike pattern) was the most effective, resulting in a 47% reduction in VALR [24].

With respect to technique modifications during running, increased knee flexion at initial contact is thought to provide better knee cushioning to dissipate impact forces [116, 122], and therefore knee-based biofeedback could be beneficial at reducing overall loading on the tibia. Although some studies measured the effects of biofeedback on knee kinematics [246–248], surprisingly, no studies were found which used thigh/knee joint based biofeedback directly to reduce loading variables during running. More studies are needed to examine the effect of providing direct technique related biofeedback on loading variables during running.

Table 4.3: Selection of studies examining the use of technique based real-time biofeedback during running.

Author	Year	Mode of Feedback	No. of Subjects	Training Environment	Feedback Parameter	Outcome Measure	Loading Results
Miller et al. [249]	2021	Auditory, Verbal, Visual	9	Treadmill	FSP, Step rate	VIP VALR Step rate Foot-strike pattern	No change -52%* +6%* 100% transition from RFS to NRFS
Hunter et al. [245]	2014	Auditory, Visual	1	Treadmill	Pelvic Kinematics	Pain Score Running Time	50% improved 12.5%* improved
Argunsah and Yalcin [250]	2020	Visual, Haptic	24	Treadmill	Gait Symmetry, Balance control	Gait Symmetry Center of mass Heart rate	More stable than no feedback* More stable than no feedback More stable than no feedback
Willy et al. [248]	2016	Visual	30	Treadmill	Step rate	VALR VILR Step rate Hip Adduction angle Eccentric Knee work per stance Eccentric Knee work per km	-19%* -18%* +9%* +3%* -27%* -21%*
Huang et al. [244]	2019	Auditory, Visual	19	Treadmill	Foot-strike, Step rate, Trunk kinematics	VIP, VALR, VILR, PTA VIP, VALR, VILR, PTA VIP, VALR, VILR, PTA	$FSP_{biofeedback}$: No change, -42%*, -17%*, -8% Step rate$_{biofeedback}$: no change, -8%, no change, -13% Anterior Trunk Lean$_{biofeedback}$: +5%, +21%*, +24%*, +11%
Roper et al. [247]	2016	Visual	16, Cont: 8	Treadmill	Foot-strike	Knee ROM Knee Abduction Knee pain score	increased flexion* reduced abduction* Pain score improved*
Chueng and Davis [224]	2011	Auditory	3	Treadmill	Foot-strike	VIP VALR VILR Foot-strike	general reduction -(15-35%) -(14-32%) (90% of footfalls converted to NRFS)
Noehren et al. [225]	2011	Visual	10	Treadmill	Hip and Pelvis Kinematics	VIP VALR VILR Hip Internal Rotation Hip Adduction angle Contra. Pelvic Drop	-10% -20%* -18%* -23%* -23%* -24%*
Willy et al. [246]	2012	Verbal, Visual	10	Treadmill	Mirror view of lower-body	Thigh adduction angle Hip Internal Rotation Hip Adduction angle Contra. Pelvic Drop	-27%* -17%* -29%* -21%*
Lorenzoni et al. [251]	2018	Auditory	13	Treadmill	Step rate	Step rate	-7%
Baggaley et al. [24]	2016	Visual	32	Treadmill	VALR, Foot-strike and step length	VALR VALR VALR	$VALR_{biofeedback}$: -31%* $FSP_{biofeedback}$: -47%* Step Length$_{biofeedback}$: -15%*
Garofolini et al. [230]	2020	Visual	20	Treadmill	VALR, Foot-strike and step rate	VALR VALR VALR VALR VALR	$VALR_{init}$: -35% $VALR_{PostKP}$: -55% Foot-strike: -60% Step rate: -30%

: Significant finding, Cont: Number of Controls, Reduced - Significantly reduced but no values provided, RFS: Rear foot-strike pattern, NRFS: Non-rear foot-strike pattern, $VALR_{init}$: Initial VALR biofeedback, $VALR_{PostKP}$: VALR biofeedback directly after technique biofeedback, Contra. Pelvic Drop: Contralateral Pelvic Drop, ROM: Range of Motion.

Difficulties with Existing Running Biofeedback Research

There are clear benefits for runners if the use of biofeedback for running re-training became more common in clinical practice, however, it has been reported that it is infrequently used in clinical settings [252]. This could be due to the fact that the vast majority of biofeedback studies utilise expensive equipment in a constrained laboratory setting with limited access [20, 23, 104, 229]. In addition, another challenge with laboratory-based solutions is the failure to replicate a runner's natural running environment. Some studies have highlighted the need to examine whether loading reductions from running retraining provided visually in a laboratory can be transferred to overground running [235, 238]. This is supported by Zhang et al. [238] who found that their treadmill based running retraining program was transferable to treadmill slope running and overground running, but not transferable to overground slope running. The need to check whether running retraining programs are transferable to overground or outdoor running could be removed if future studies focused on providing running retraining (both training and intervention evaluation) in these environments. A recent review paper by Van Gelder et al. [25] found that only 4% of the reviewed papers (8/173) provided biofeedback in a non-laboratory environment. This suggests that there should be a focus on portable/mobile-based running retraining solutions which would be free from location constraints. In addition, while most biofeedback studies (where feedback is provided in a laboratory) need to provide phased feedback in order to generate retention for transference to outdoor running, a focus on mobile provision of biofeedback would remove the need for phased feedback and instead (with the support of a clinician) could provide a program of intermittent feedback provided on a scheduled basis i.e., once a week, once a fortnight etc. This would still support technique retention and avoid over reliance on the feedback itself. Future focus should remain on the use of small portable devices such as mobile phones would allow the provision of auditory biofeedback in any environment, and inertial sensors which provide location independent, low-cost data capture of variables related to loading and technique [253].

Chapter 5

Related Work: Machine Learning

5.1 Introduction to Machine Learning

Living in the era of big data, access to data covering a range of aspects of everyday life has become more readily available [254]. With this, the scope for problems which can be explored and solved with data mining are endless [255, 256]. Machine learning is the process of using data as an input for algorithms to build models which can learn patterns from the data. The learned patterns can be used in order to make future predictions/decisions on similar unseen data to fulfil a particular objective [257]. Machine Learning is a subset of artificial intelligence which includes Traditional Machine Learning and Representational Machine Learning (commonly referred to as Deep Learning) techniques. Machine learning analysis of a problem may succeed where human analyses fail in attempting to reveal underlying patterns in data that may not be explicit.

Each sample instance within a dataset involved in model training or model testing is described by a set of features. These features are the variables which are used to describe the problem and to discover underlying relationships between feature values and output prediction. Features can be discrete, continuous or categorical values. Good features can be chosen by an expert with domain knowledge with some intuition on the data that may be useful for prediction. This is one of the main differences between traditional machine learning approaches and deep learning approaches, using traditional algorithms features must be chosen and extracted by the researcher, whereas in deep learning the models themselves automatically extract features to describe the problem during the training phase [256]. This thesis focuses on traditional algorithms (non-deep learning) only.

Broadly there are three main types of machine learning algorithms: supervised learning, unsupervised learning and reinforcement learning [256]. In supervised learning, sets of features are used as inputs to the learning models with a known label associated with each feature set. These labels are used in model training to associate combinations of feature values with an output label

[257]. This differs from unsupervised learning where known labels are not involved in the model training process, and instead class separation is performed by other means e.g., clustering [255, 257]. Reinforcement learning works by reward-based learning; it aims to make decisions which lead to maximising the reward score by trialling many iterations of possible decisions within the model [255, 257].
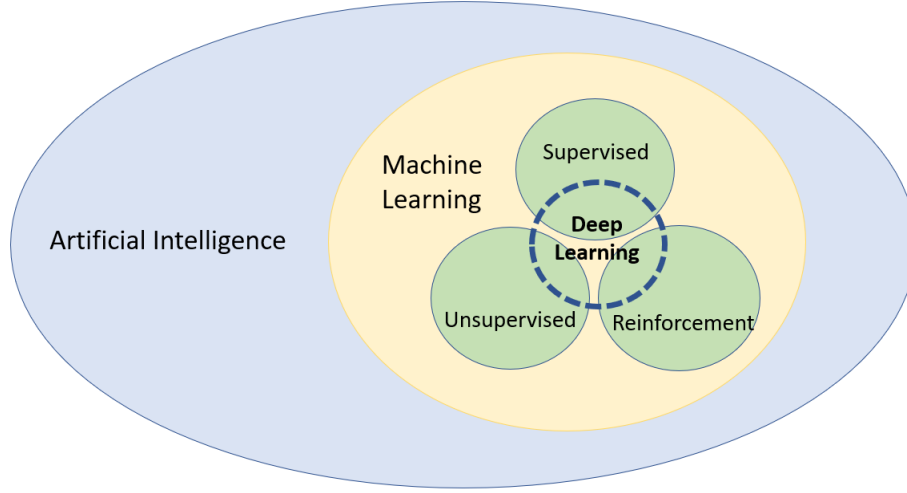
Figure 5.1: The relationships between the different elements of Machine Learning described above. Adapted from [3].

The research outlined in this thesis examines supervised learning only, this decision was made as it was important to the research that any positive outcomes (e.g., potentially predicting likelihood of injury in runners) had a level of explainability. Explainable AI is a research area of its own, for more information on explainable AI a thorough survey which discusses explainability with a focus on supervised learning can be found in Burkart and Huber [258].

5.2 Supervised Learning

With the focus on supervised learning, these types of algorithms aim to learn a model $H(X) = y$, which maps a set of features X, to a target outcome prediction y. Initially the dataset is split into a training set and testing set, where a sample instance X in each set represents a set of features parsed from the dataset (where $X = \{feature_1, feature_2, ..., feature_p\}$, for p $\in \mathbb{N}$ available features). The model is trained using tuples of feature sets and known class labels from the training set Tr (where Tr $= \{(X_1, y), (X_2, y), ..., (X_m, y)\}$, for m $\in \mathbb{N}$ sample instances in the training set) [255, 257]. On testing the model's performance, sets of features (X) from the testing set (specifically not used as part of model training) are input to the model to retrieve a target outcome prediction ($H(X) = y$).

Supervised learning problems can be analysed using classification or regression (figure 5.2). For classification problems the learning model predicts y as a categorical label (e.g., predicting running or walking, injured or not injured, etc.) and for regression, the learning model predicts y as a continuous value (e.g., predicting a joint angle). The research problem described in Section 9, to build a model capable of predicting future likelihood of incurring a running related injury, is a classification problem to differentiate between two binary classes of injured and uninjured runners. A number of commonly used supervised learning approaches for classification (machine learning classifiers) were implemented in Section 9 to examine this aim.
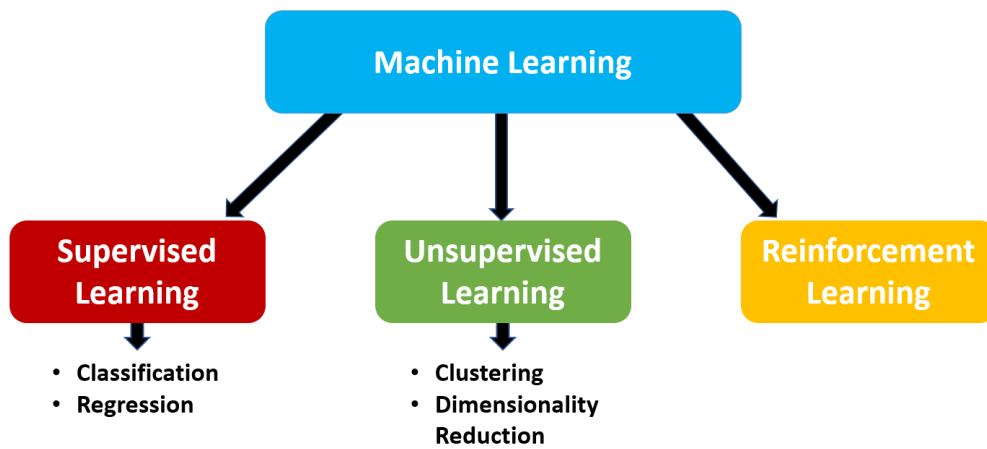
Figure 5.2: The three main types of Machine Learning algorithms.

5.2.1 Supervised Learning Classifiers

In the following section, the general mechanics of the machine learning classifiers which were used in Section 9 are explained. Figure 5.3 describes the general steps involved in all supervised learning classification approaches.

Machine Learning Terminology:

- **Feature Selection/Feature Reduction**: This involves performing techniques on the full feature set in order to identify features which may be more informative and important in the classification process, and to exclude those features which are deemed to be less relevant.

- **Hyperparameters**: These are model parameters used within the classifiers to direct the learning process. Varying the value of hyperparameters can impact the accuracy of the classification model.

- **Hyperparameter tuning**: The use of cross validation (validation samples not involved in model training) to test a range of values for certain hyperparameters, in order to select the parameter values which result in the best classification of validation samples.

- **Overfitting**: Overfitting describes when the model remembers the training data too well in that it loses the ability to generalise to unseen data.

- **Underfitting**: Overfitting describes when the model is unable to learn from the training data or generalise to unseen test data. In essence the model has failed to learn any relationship between the input features and target prediction labels.
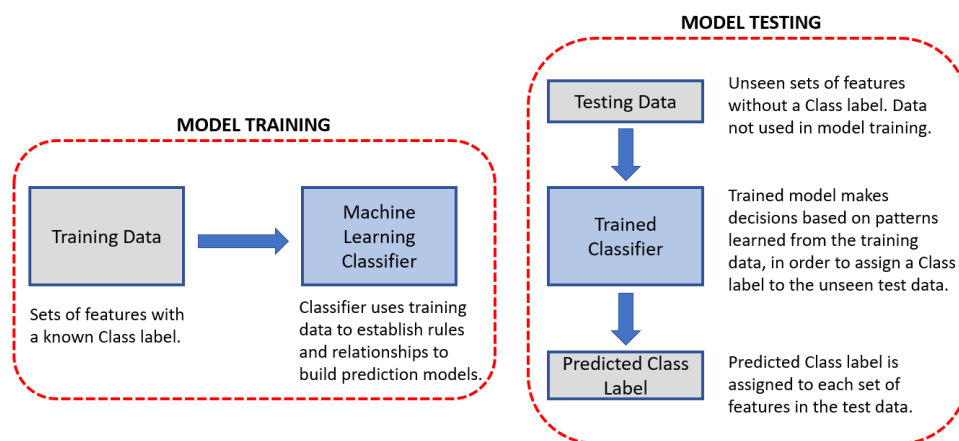
Figure 5.3: The main steps involved in training and testing a supervised classification model.

Naive Bayes

The structure of the Naive Bayes Classifier is founded on Bayes' Theorem. Bayes' theorem is a probabilistic model which calculates the probability of an event given the prior probability of other

related events occurring. Naive Bayes calculates the posterior probability (the probability of event A occurring given that event B has occurred) by combining known prior probability and a measure of likelihood. Equation 5.1 shows the formula for Bayes' Theorem. This classifier has far fewer hyperparameters to tune during the training process (zero or one parameter) compared to other supervised classification algorithms, for example decision trees which can have >5 hyperparameters. This aspect allows for faster model training times and makes this classifier easily applicable to large data sets. In terms of explainability, this model returns the probabilities of predicted classes and so it can be applied to many machine learning problems.

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)} \tag{5.1}$$

Where:

P(A | B) : Posterior probability of the class when predictor is given

P(A) : Prior probability of the class

P(B | A) : Probability of the predictor when class is given

P(B) : Prior probability of the predictor

Logistic Regression

Logistic Regression is a supervised classifier similar to ordinary linear regression where usually the prediction separates samples between two distinct classes. This class separation is based on the logistic function (also called the Sigmoid function), which is a S-shaped curved that maps values to a prediction between 0 and 1. The value predicted between 0 and 1 on the logistic function represents the probability of a sample belonging to one class or the other. A threshold exists i.e., 0.5 for binary classification, where mapped values above 0.5 are assigned to one class and mapped values below 0.5 are assigned to the other class. Logistic Regression can be generalised for the multi-class classification case where prediction is performed on >2 classes. Only binary classification is performed in the work in this thesis.

K-Nearest Neighbour

K-Nearest Neighbour classifier is one of easiest supervised learning approaches to both implement and understand. The underlying idea of this classifier is that when each sample is described by a point
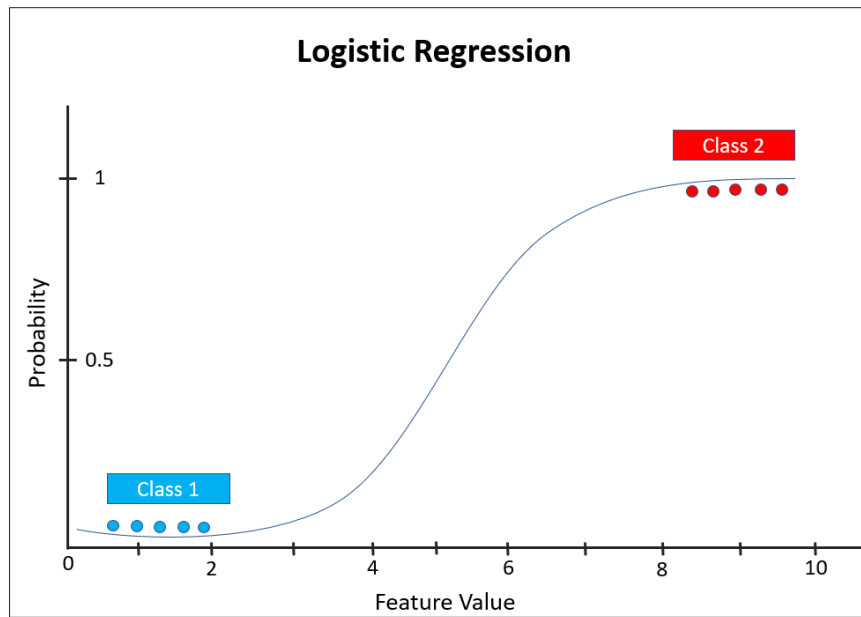
Figure 5.4: The Logistic function used to map feature values to classes in the range of 0 to 1.

in an N-dimensional space, similar samples are positioned closer together. For each unlabelled test sample, K-nearest Neighbour searches the entire set of training data for the K (where $K \in \mathbb{N}$) nearest training samples to the unlabelled test sample using some fixed distance metric. The idea is that these K chosen samples are most similar to the unlabelled sample. The K-nearest samples are used to assign a class label to the unlabelled sample by a majority vote of the K known labels. A distance metric is chosen to reduce the distance between related samples, and to maximise the distance between unrelated samples. Varying the value of K can result in different classification performance for any given problem, and so usually K is chosen as a hyperparameter via cross-validation techniques. This classifier can be particularly hindered by unimportant features for classification, as each irrelevant feature increases the size of the KNN feature space and increases the likelihood of nearest neighbours not being related to the correct classification. Typically, some feature selection or feature reduction techniques are explored before the use of supervised learning classifiers, in order to reduce the feature set down to fewer, more informative features.

Decision Trees

Decision tree classifiers build a structured tree model from the labelled training data. The structure of the tree model is defined by the feature values of each sample in the training data, where new unlabelled samples are assigned a class label based on their feature values. The top node of the tree (root node) is defined by the feature that best splits the training data. It has been shown that no single
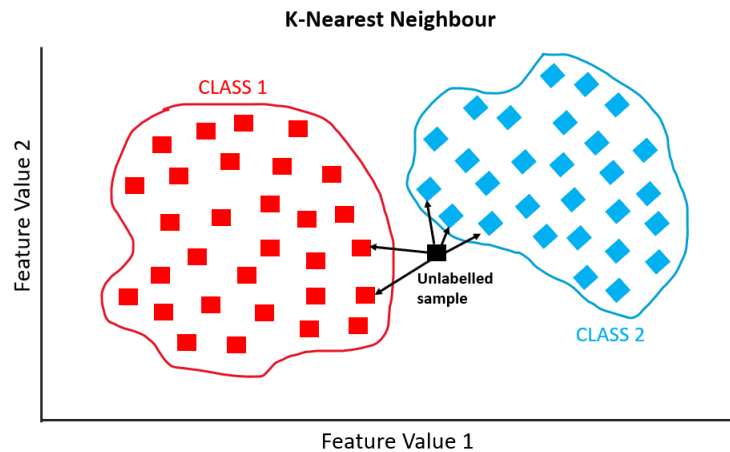
Figure 5.5: The K-Nearest Neighbour classifier using the k=5 nearest labelled samples to assign a class to the unlabelled sample.

best approach exists to identify the best feature for the root node, and so a number of techniques are available to examine its selection. Following the division of the data at the root node, each following node represents a different feature where the tree's branches off this node represent the values that the node feature may possess (e.g., node with binary feature, branch one: yes, branch two: no). The feature values are sorted in this way for each new unlabelled sample until the tree reaches a final leaf node which dictates a class label to assign the sample. In order to reduce the risk of overfitting, pruning techniques can be employed which involve removing unhelpful sub-sections of the decision tree model. One great advantage of decision trees is the explainability of the classification decisions. Simply following the decisions made at each node down the tree leads to the final classification label.

Ensemble Classifiers

Ensemble classifier as the name suggests are collections of a number of classifiers. This may be a single model made up of multiple instances of a single classifier (e.g., decision trees) or made up of a number of different classifiers (ensemble of decision tree and naive bayes classifiers). Research has shown that ensemble methods typically perform better classification than using a single classifier alone (one decision tree classifier). In theory, ensemble classifiers create a single stronger classifying model by using a set of weaker learners. These ensemble classifiers are either referred to as boosting or bagging classifier depending on how the training data is used. A boosting ensemble classifier describes the approach where the classifiers within the ensemble are trained iteratively on the full training set, where each ensemble learns from the misclassifications of the previous ensemble and adjusts the classification model accordingly to improve overall accuracy. A bagging ensemble classifier describes

the approach where each sub classifier within the ensemble is trained using a different random subset (random sampling with replacement) of the entire training set.
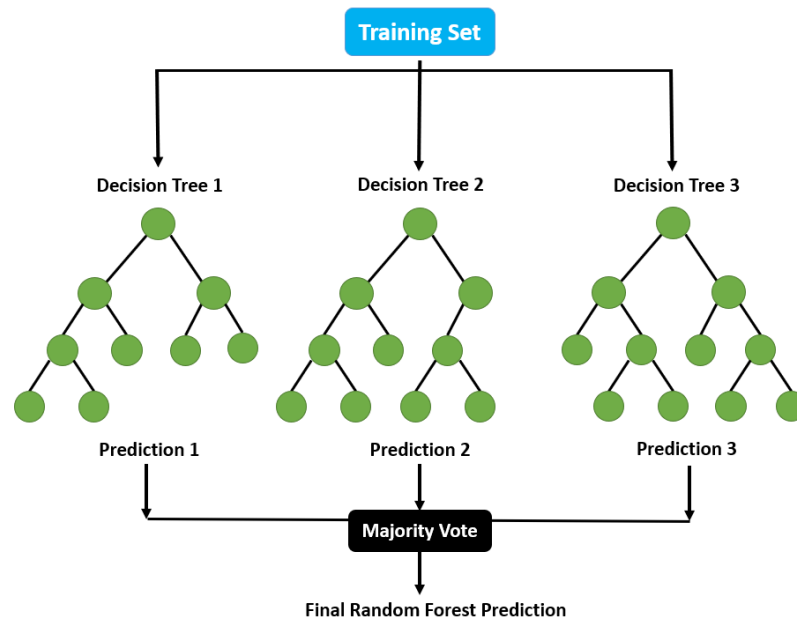
Figure 5.6: A Random Forest ensemble classifier made up of three decision trees. Each decision tree is trained independently on a subset of the full training data. The majority outcome prediction is used to assign classes to unlabelled samples.

Random Forest

Random forest is an example of an ensemble classifier. A random forest classifier is a single model made up of many decision tree classifiers used to make a prediction. As it is a bagging ensemble, each decision tree within the random forest is created using a subset of the full training data which is sampled with replacement. The same single sample may appear in multiple decision trees, or in none, due to this sampling approach, but many combinations of the training data will be covered. Each decision tree is trained as an independent classifier. The final classification of an unlabelled sample is performed by passing the sample to each decision tree within the ensemble and taking a majority vote of the resulting classification scores from each tree.

Support Vector Machine

Support Vector Machine (SVM) classifiers can be applied to a linear or non-linearly separable classification problem. The broad aim of this classifier is to define a decision boundary in n-dimensional space which separates the samples of the prediction classes. In the linear case, the

classifier attempts to create a hyper-plane (decision boundary) between the samples of unique classes, which maximises the distance between the hyper-plane and the closest sample instances of either class (the margin). The closest sample instances are known as support vectors, and they represent two parallel margins of which the distance to the hyper-plane is to be maximised. These margins can be referred to as soft or hard margins, depending on the willingness of the model to accept misclassifications in return for a more stable model (less sensitive to noise). Slack variables included in the modelling control the degree of penalty for misclassifications, and therefore the severity of the decision boundary. For non-linear classification problems, where a hyper-plane cannot simply split the class decision boundary, a kernel function is used to transform the data to a higher dimensional feature space. The idea is that in a higher dimensional space it may be easier to separate the classes. Traditionally SVM is implemented as a single classifier but in this thesis SVM was implemented as a bagging ensemble in order to reduce the computational overhead that was discovered using a large number of training features.
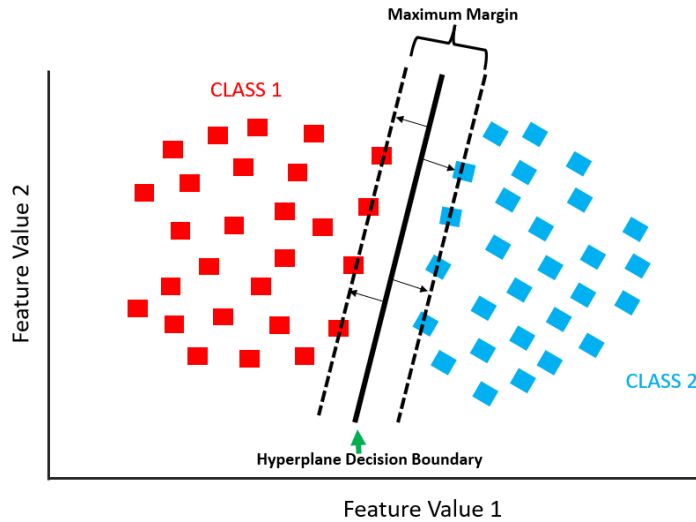
Figure 5.7: Support Vector Machine classifier to solve a linearly separable problem. The distance between the margin boundaries (defined by samples of unique classes) are maximised in order to create a hyper-plane decision boundary which separates unique classes.

5.3 Applications of Machine Learning with Movement Technique

Given the focus of one of the main studies of this PhD, initially the primary aim of this section of the review was to highlight existing research involving the use of machine learning to predict running related injuries based on movement technique from inertial sensor and/or kinematic data. However,
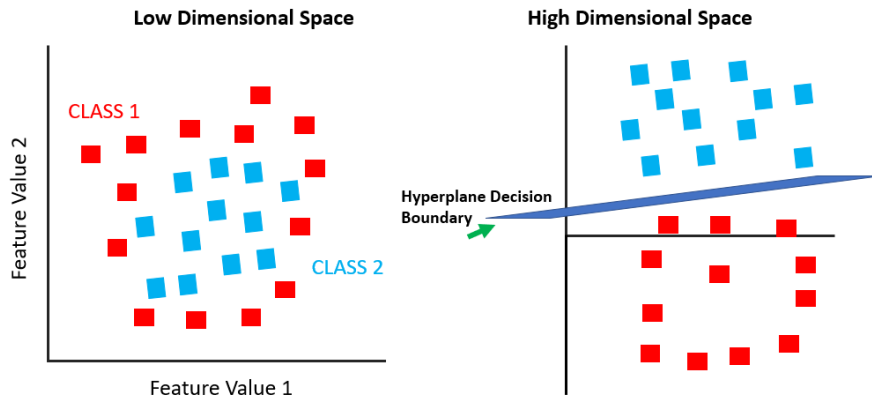
Figure 5.8: Support Vector Machine classifier to solve a non-linearly separable problem. Training samples are transformed to a higher dimensional feature space using a kernel function. Transformation of the data using a kernel function allows the creation of a model which was not linearly separable in the lower dimensional space, to define a hyper-plane decision boundary in the higher dimensional space.

given that there are so few studies in this focused area, this review of literature also examines the use of machine learning and inertial sensor/kinematic data for the prediction of: (ii) non-running injuries, (iii) diseases associated with movement impairment, (iv) activity recognition and (v) a number of non-injury variables predicted during walking and running. In terms of the range model strength, AUC values of 0.6-0.7 were considered poor, 0.7-0.8 fair, 0.8-0.9 good, and >0.9 excellent [259–261]

5.3.1 Prediction of Running Injuries

Seventeen previous research studies have been identified which were relevant to review for injury prediction using machine learning and technique-based data (inertial sensors and/or kinematic data). These studies can be grouped based on the type of injuries which were examined: either specific or general running related injuries (see tables 5.1 and 5.2, respectively).

With regard to specific injuries, 10 studies were identified which investigate the prediction of: hamstring strain injuries [262], Anterior Cruciate Ligament (ACL) injuries [263–269], knee and ankle injuries [270] and brachial plexus injury [271]. Only one study [268] used data collected during running gait to train their prediction model, with the remaining studies reviewing a variety of movements (walking, jogging, jump landing). Each of the specific injuries examined in these non-running studies (excluding brachial plexus injury) could be typically incurred during running activities. Overall, 10 unique classifiers were used across the 10 studies, in order of most evaluated these classifiers were: Support Vector Machine (SVM), Neural Networks, Decision Tree, Logistic Regression, Random Forest, K-Nearest Neighbour (KNN), Ensemble methods, Naive Bayes, Linear Discriminant analysis

and Stacking. SVM and Neural networks were the most used classifiers, appearing 5 times each across these 10 studies.

While it is difficult to compare across studies, the majority of the reviewed studies evaluated their classifier using a measure of standard accuracy (number of correct predictions divided by total number of predictions) or AUC (area under the receiver operating characteristic curve). On average across the 10 studies the range of accuracies for each classifier were: SVM (0.46 AUC, 59-97% accuracy), Neural Networks (0.68-0.75 AUC, 73-96% accuracy), Random Forest (0.63-0.89 AUC, 80% accuracy), Decision Tree (0.6-0.84 AUC), ensemble (72-90% accuracy), KNN (53-85% accuracy), Naive Bayes (72% accuracy), Logistic Regression (0.57-0.68 AUC, 77% accuracy), Linear Discriminant Analysis (68% accuracy) and stacking (73% accuracy).

Overall, across the 10 studies the level of prediction accuracy was high and typically the best classifier of each study performed > 85% standard accuracy or >0.75 AUC. However, Jauhiainen et al. [270] reported a weak level of prediction (AUC 0.63) using random forest to identify knee and ankle injuries in young athletes. This suggests perhaps the features/variables for prediction in Nazarahari et al. [271] and Ayala et al. [262] acted as better predictors for these specific injuries (brachial plexus injury: 90% accuracy and hamstring strain injuries: 0.84 AUC, respectively). Tedesco et al. [268] was the only study to specifically examine running gait, and although the reported accuracy of the model is rather low (73% accuracy), their study showed that differences in running gait between healthy and post-ACL gait patterns could be detected even a number of years after the injury occurred.

With regard to general injuries, 7 studies were identified which investigate the prediction of running injuries in recreational runners [272], triathletes [273], student athletes [274] and team sports athletes (soccer [275–277], futsal [278]). Overall, 5 unique classifiers were used across the 7 studies, in order of most evaluated these classifiers were: Ensemble methods, Random forest, Logistic Regression, Decision trees and Naive Bayes. Random Forest was the most used classifier appearing 4 times across these 7 studies.

All of the reviewed studies examining general running injury studies evaluated their classifier using a measure of AUC. On average across the 7 studies the range of AUC for each classifier were: Ensemble methods (0.51-0.75 AUC), Random forest (0.69 - 0.8 AUC), Logistic Regression (0.59-0.65 AUC), Decision trees (0.76 AUC) and Naive Bayes (0.6 AUC). The most accurate model was outlined in Martinez et al. [273], where a random forest model had a predictive score of running injuries of approximately 0.8 AUC. This study was limited in participant numbers (19 participants) which

calls into question the generalizability of the findings. [279] includes a comparison of the prediction of general injury and specifically hamstring strain injury and reported higher accuracies associated with the specific injury prediction. This finding in combination with the good performance of specific injury prediction described above may indicate that identification of specific injuries may be easier than general injuries where predictors may be conflicting.

It is important to note the use of AUC as a metric for prediction across the majority of studies investigating injury prediction. This metric is particularly useful in the event of unbalanced classes within datasets [279], which is an accepted problem in most injury prediction studies where the number of injuries are typically far less than the numbers of non-injured controls. AUC provides a more accurate estimation of model performance than standard accuracy measures can provide, which is obvious if you consider a classifier that simply guesses the same class (the majority class) for all test samples and produces a false inflated model accuracy [280].

$$\text{Accuracy } = \frac{\text{TN} + \text{TP}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{5.2}$$

Across all 19 studies identified which examine the prediction of injuries, the most used classifiers were: random forest, logistic regression, ensemble methods, decision trees, SVM, neural networks, K-NN and Naive Bayes, all of which were examined in more than one study. These classifiers will form the basis of the methodology explored in Section 9 for the prediction of running related injuries.

Table 5.1: Selection of studies examining the use of machine learning and technique data for the prediction of specific injuries.

Author	Year	Study Objective	Relevant Data Type	Movement Captured	Number of Subjects	Validation Method	Classifier Used	Prediction Accuracy
Mandalapu et al. [263]	2019	Prediction of ACL injury	IMU	Walking, jogging	131	5-Fold CV	DT	0.60 AUC
							LogReg	0.68% AUC
							MLP	0.75 AUC
							NN	0.68 AUC
							RF	0.89 AUC
							SVM	0.46 AUC
Richter et al. [264]	2021	Prediction of ACL injury	Kinematics	Jump	822	-	LogReg	0.57 AUC
Borgstrom et al. [265]	2015	Prediction of ACL injury	IMU	Knee movements	61	LOSO	SVM	97% Accuracy
Wu et al. [266]	2018	Prediction of ACL injury	Kinematics	Walking	46	LOSO	NN	96% Accuracy
Richter et al. [267]	2019	Prediction of ACL injury	Kinematics	Jumping, change of direction	218	stratified shuffle split	DT	65% Accuracy
							KNN	53% Accuracy
							LDA	68% Accuracy
							LogReg	77% Accuracy
							NN	80% Accuracy
							SVM	59% Accuracy
							RF	80% Accuracy
Tedesco et al. [268]	2020	Prediction of ACL injury	IMU	Running, Side-stepping	12	LOSO	GB	72% Accuracy
							KNN	72% Accuracy
							MLP	73% Accuracy
							NB	72% Accuracy
							Stacking	73% Accuracy
							SVM	71% Accuracy
Taborri et al. [269]	2021	Predict ACL injury risk	MIMU	Jump landing	39	9-fold CV	DT	90% Accuracy
							KNN	85% Accuracy
							SVM	95% Accuracy
Jauhiainen et al. [270]	2020	Prediction of severe knee and ankle injuries	Kinematics	Vertical drop landing	314	10-fold CV	LogReg	0.65 AUC
							RF	0.63 AUC
Ayala et al. [262]	2019	Prediction of hamstring strain injury	Kinematics	ROM tests	96	3-fold CV	DT	0.84 AUC
Nazarahari et al. [271]	2021	Predict Brachial Plexus injury	IMU	Shoulder movements	30	10-fold CV	Ensemble	90% Accuracy

Std. Deviation: Standard Deviation, RateNorm: Rate of acceleration for normal running, RateBio: Rate of acceleration for running with biofeedback.

Table 5.2: Selection of studies examining the use of machine learning and technique data for the prediction of general running related injuries.

Author	Year	Study Objective	Relevant Data Type	Movement Captured	Number of Subjects	Validation Method	Classifier Used	Prediction Accuracy
Rossi et al. [275]	2017	injury risk in football players	IMU	Soccer training	26	-	DT RF	0.76 AUC 0.71 AUC
Martínez-Gramage et al. [273]	2020	Predicting running related injuries	IMU	Running	19	10-fold CV	RF	0.8 AUC
López-Valenciano et al. [276]	2018	Predicting running related injuries	Kinematics	Soccer training	132	5-fold CV	Ensemble	0.75 AUC
Gore [272]	2020	Predicting running related injuries	Kinematics	Running	282		Adaboost Bagged SVM Ensemble LogReg NB RF	0.63 0.51 AUC 0.7 AUC 0.59 AUC 0.6 AUC 0.71 AUC
Oliver et al. [277]	2020	Prediction of soccer injuries	Kinematics	Jump movements	355	5-Fold CV	Ensemble Univariate LogReg Multivariate LogReg	0.66 AUC 0.65 AUC 0.65 AUC
Henriquez et al. [274]	2020	Predict LE-Musculoskeletal Injury	Kinematics	ROM tests	122	5-fold CV	RF	0.69 AUC
Ruiz-Pérez et al. [278]	2021	Predict LE-Musculoskeletal Injury	Kinematics	ROM tests	139	5-fold CV	DT	0.76 AUC

LE-Musculoskeletal:Lower Extremity Musculoskeletal.

81

5.3.2 Prediction of Diseases

Machine learning has proven to be a potentially useful tool in disease prediction [281]. Many studies have examined the use of non-technique data with machine learning to predict various diseases e.g., heart disease, breast cancer (a wide selection of this work can be found in: Uddin et al. [281]). However, this review solely focuses on a selection of those which have used only inertial sensor and/or kinematic data to measure movement technique for the prediction of diseases related to motor impairment (i.e., where differences in movement related to pathologies can be measured). A number of studies have examined the use of movement technique and machine learning to predict different diseases associated with movement impairment, of which we have included 25 studies in this review (see tables 5.3 and 5.4) to give a scope of the application area for disease prediction.

With regard to specific injuries, the selected 25 studies investigate the prediction of: Parkinson's disease [282, 283], Alzheimer's disease [284–290], Supranuclear Palsy [291], Cerebral Palsy [292], Huntington's disease [293], Stroke [293, 294], Scoliosis [295], Hip Osteoarthritis [296], Hip Arthroplasty [297], Elbow Spasticity [298] and Friedreich's Ataxia [299]. The majority of previous research in this area seems to investigate the identification of symptoms related to Parkinson's disease [300]. This majority was also reflected in the selection of studies in this review with 12 of 25 studies examining some characteristic of Parkinson's disease. Overall, 11 unique classifiers were used across the 25 studies, in order of most evaluated these classifiers were: Support Vector machine (SVM), Random Forest, Ensemble, K-Nearest Neighbour (KNN), Neural Network (including multi-layer perceptron and radial basis function neural networks), Decision tree, Naive Bayes, Logistic Regression, Linear Discriminant Analysis, Deep Neural Networks (including convolutional neural network, deep belief network, long short-term memory network) and Linear Regression. SVM and Random Forest were the most used classifiers, appearing 20 times and 9 times across these 25 studies, respectively.

While it is difficult to compare across studies, the majority of the reviewed studies evaluated their classifier using a measure of standard accuracy (the number of correct predictions divided by the total number of predictions). On average across the 25 studies the range of accuracies for each classifier were: SVM (66-95%), Random Forest (84-97%), Ensemble (72-96%), KNN (67-90%), Neural Network (85-100%), Decision tree (62-86%), Naive Bayes (73-97%), Logistic Regression (77-90%), Linear Discriminant Analysis (65-85%), Deep Neural Networks (97-99%) and Linear Regression (94%).

Overall, across the 25 studies the level of prediction accuracy was high and typically the best

classifier of each study performed $> 86\%$. Three studies which had lower levels of performance for their best classifier had accuracies of 78-81%, which is acceptable. All of the diseases examined interact with motor function and lead to movement impairment whether this is abnormalities in gait or movement. These abnormalities created clear predictors in the captured movement technique (kinematics or IMU signals) which allowed for excellent levels of classification across different studies and across various classifiers.

Table 5.3: Continued: Selection of studies examining the use of machine learning and inertial sensors/kinematic data for the prediction of disease.

Author	Year	Study Objective	Relevant Data Type	Movement Captured	Number of Subjects	Validation Method	Classifier Used	Prediction Accuracy
Ghaderyan et al. [284]	2018	Detection of Alzheimer's disease	Kinematics	Handwriting tasks	43	5-Fold CV	Linear Regression	94% Accuracy
Costa et al. [285]	2016	Detection of Alzheimer's disease	IMU	Postural tasks	72	3-Fold CV	DBN MLP RBFNN SVM	97% Accuracy 97% Accuracy 97% Accuracy 91% Accuracy
Filtjens et al. [282]	2020	Detection of Parkinson's symptoms	Kinematics	Walking and turning	15	LOSO CV	CNN LSTM	99% Accuracy 99% Accuracy
Lamba et al. [283]	2021	Detection of Parkinson's disease	Kinematics	Handwriting tasks	77	10-Fold CV	Adaboost RF SVM XGBoost	96% Accuracy 94% Accuracy 94% Accuracy 94% Accuracy
Aich et al. [286]	2018	Detection of Parkinson's disease	Kinematics	Walking	40	-	NB RF SVM	97% Accuracy 97% Accuracy 99% Accuracy
Drotar et al. [287]	2014	Detection of Parkinson's disease	Kinematics	Handwriting tasks	75	LOSO	SVM	86% Accuracy
Duque et al. [288]	2020	Distinguish between patients with Parkinson's Disease and Essential Tremor	Gyroscope	arm poses	41	random splits	Ensemble KNN LogReg NB SVM	76% Accuracy 76% Accuracy 77% Accuracy 73% Accuracy 78% Accuracy
Tien et al. [289]	2010	Detection of Parkinson's disease	IMU	Walking	39	10-Fold CV	SVM	98% Precision
Moon et al. [290]	2020	Distinguish between patients with Parkinson's Disease and Essential Tremor	IMU	Walking	567	3-Fold CV	DT GB KNN LogReg NN RF SVM	0.53 F1-score 0.59 F1-score 0.49 F1-score 0.53 F1-score 0.61 F1-score 0.56 F1-score 0.55 F1-score
Drotar et al. [301]	2016	Detection of Parkinson's disease	Kinematics	Handwriting Task	75	10-Fold CV	Adaboost KNN SVM	72% Accuracy 79% Accuracy 81% Accuracy
de Araújo et al. [302]	2020	Detection of Parkinson's disease	IMU	Resting hand tremors	50	10-Fold CV	DT GB KNN LDA LogReg RF SVM	85% Accuracy 75% Accuracy 90% Accuracy 85% Accuracy 90% Accuracy 90% Accuracy 75% Accuracy

Std. Deviation: Standard Deviation, RateNorm: Rate of acceleration for normal running, RateBio: Rate of acceleration for running with biofeedback.

Table 5.4: **Selection of studies examining the use of machine learning and inertial sensors/kinematic data for the prediction of disease.**

Author	Year	Study Objective	Relevant Data Type	Movement Captured	Number of Subjects	Validation Method	Classifier Used	Prediction Accuracy
Jeon et al. [303]	2017	Classification of Parkinson's Tremor Severity	IMU	Resting hand tremors	85	LOSO CV	DT KNN LDA SVM	86% Accuracy 81% Accuracy 81% Accuracy 92% Accuracy
Caramia et al. [304]	2018	Detection of Parkinson's disease	IMU	Walking	50	5-Fold CV	DT KNN LDA NB SVM	62% Accuracy 67% Accuracy 65% Accuracy 78% Accuracy 66% Accuracy
Rovini et al. [261]	2018	Detection of Parkinson's disease	IMU	Walking and Foot postures	90	10-Fold CV	NB RF SVM	95% Accuracy 97% Accuracy 95% Accuracy
Park et al. [305]	2021	Detection of Parkinson's symptoms	IMU	hand movements	46	5-Fold CV	LogReg NN	0.8 AUC 0.9 AUC
Printy et al. [306]	2014	Classification of Motor Impairment Severity in Parkinson's Disease	IMU	Hand movements	26	6-Fold CV	RF SVM	95% Accuracy 89% Accuracy
De Vos et al. [291]	2020	Discriminating progressive Supranuclear Palsy from Parkinson's disease	IMU	Walking	60	10-Fold CV	LogReg RF	80% Accuracy 88% Accuracy
Zhang and Ma [292]	2019	Classification of gait patterns in children with Cerebral Palsy	Kinematics	Walking	200	10-Fold CV	DT KNN LDA NB NN RF SVM	84% Accuracy 78% Accuracy 84% Accuracy 82% Accuracy 94% Accuracy 84% Accuracy 85% Accuracy
Park et al. [294]	2020	Classification of Stroke Severity	Accelerometer	Arm/Leg movements	17	5-Fold CV	Ensemble SVM	0.91 AUC 0.86 AUC
Mannini et al. [293]	2016	Discriminating post-stroke patients from Huntington's disease patients and healthy controls	IMU	Walking	42	LOSO CV	SVM	91% Accuracy
Cho et al. [295]	2018	Detection of Scoliosis	IMU	Walking	42	5-Fold CV	SVM	91% Accuracy
Laroche et al. [296]	2014	Detection of Hip Osteoarthritis	Kinematics	Walking	40	5-Fold CV	SVM	88% Accuracy
Teufl et al. [297]	2021	Detection of hip arthroplasty gait patterns	IMU	Walking	46	LOSO CV	SVM	0.96 AUC
LeMoyne et al. [299]	2016	Detection of Friedreich's ataxia gait patterns	IMU	Walking	2	-	MLP	100% Accuracy
Kim et al. [298]	2020	Classification of Elbow Spasticity Severity	IMU	Passive Stretch Test	48	LOSO CV	DT LDA MLP RF SVM	77% Accuracy 81% Accuracy 83% Accuracy 92% Accuracy 72% Accuracy

Std. Deviation: Standard Deviation, RateNorm: Rate of acceleration for normal running, RateBio: Rate of acceleration for running with biofeedback.

5.3.3   Human Activity Recognition

The area of human activity recognition is hugely popular in existing research exploring machine learning models with technique data (inertial and/or kinematic) [307]. Twenty-five studies have been included in this review (see tables 5.5 and 5.6) for the purpose of providing a scope of the application area for activity recognition. These 25 studies examined a range of human activities including: examining prediction of various basic movements (e.g., walking, standing) [308–314], but also specific daily living activities (e.g., drinking, eating, scrubbing) [315, 316], fall prediction in the elderly [317–323], prediction of fainting and epileptic seizures [324], examining walking conditions [325], specific job activities (firefighter [315]) and specific sport movements (walking versus jogging [322], jogging versus sprinting [326], tennis vs. golf [316]). A number of studies [327–331] rather than focusing on the correct classification of a number of different movements, focused on the classification of the correct or incorrect performance of specific movements (i.e., correct squat posture [327]). These studies take advantage of knowing common postural and movement mistakes made when performing certain movements and for the purpose of rehabilitation and movement training can provide insight into technique performance. This is essentially a different definition of the same problem which remains either a binary class (squatting correctly vs squatting incorrectly) or multi-class classification (squatting correctly vs. known squat mistake 1 vs known squat mistake 2).

Overall, 12 unique classifiers were used across the 25 studies, in order of most evaluated these classifiers were: Support Vector Machine, Decision tree, Random Forest, K-Nearest Neighbour, Neural Network, Naive Bayes, Ensemble, Deep Neural Network, Logistic Regression, Linear Discriminant Analysis, Hidden Markov Model and Dynamic Time Warping. SVM and decision trees were the most used classifiers, appearing 15 times and 13 times across these 25 studies, respectively.

The majority of the reviewed studies evaluated their classifier using a measure of standard accuracy with the exception of two studies which used area under the ROC Curve (AUC). On average across the 25 studies the range of accuracies for each classifier were: Support Vector Machine (72-99% Accuracy; 0.86-0.97 AUC), Decision tree (84-97%), Random Forest (61-96%; 0.89 AUC), K-Nearest Neighbour (62-99%), Neural Network (83-99%; MLP 0.90 AUC), Naive Bayes (82-99%), Ensemble (62-95%), Deep Neural Network (89-99%; LSTM 0.94 AUC), Logistic Regression (94-97%), Linear Discriminant Analysis (72-84%), Hidden Markov Model (0.91 AUC) and Dynamic Time Warping (98%). SVM and decision trees were the most used classifiers, appearing 15 times and 13 times across these 25 studies, respectively.

Overall, across the 25 studies a range of classifiers have been shown to perform extremely well, with the majority of studies included having reported a best classification accuracy of approximately >=90% (see tables 5.5 and 5.6 for details). The excellent levels of high classification accuracy could be attributed to the comparatively different signals which described between most of the activities examined (e.g., standing vs. walking vs. jumping etc.), which allowed for easy differentiation between activities. With this in mind, some studies aimed to classify movements with little obvious difference in movement (e.g., standing on two legs with open eyes vs. standing on two legs with closed eyes) and so, more difficult to uniquely classify [314, 332]. These studies still achieved high accuracies of >95% [314, 332].

Table 5.5: Selection of studies examining the use of machine learning and technique data for the prediction of human activity recognition.

Author	Year	Study Objective	Relevant Data Type	Movement Captured	Number of Subjects	Validation Method	Classifier Used	Prediction Accuracy
O'Reilly et al. [330]	2017	Classification of correct movements performed	MIMU	Deadlift	80	LOSO CV	RF	63-75% Accuracy
Taylor et al. [331]	2010	Classification of correct movements performed	Accelerometer		9	LOSO CV	AdaBoost	62-80% Accuracy
Hassan et al. [310]	2018	Classification of Daily Life Activities	IMU	ADL	-	-	ANN DBN SVM	89% Accuracy 96% Accuracy 94% Accuracy
Lau et al. [325]	2008	Classification of walking conditions	IMU	Walking	3	LOSO CV	ANN BBN RBFNN SVM	99% Accuracy 94% Accuracy 97% Accuracy 100% Accuracy
Barshan and Yüksek [333]	2013		MIMU	ADL	8	10-fold CV	ANN DT NB	99% Accuracy 95% Accuracy 94% Accuracy
Minarno et al. [314]	2020		IMU	ADL	30	-	DT KNN XGB LogReg RF SVM	86% Accuracy 90% Accuracy 94% Accuracy 96% Accuracy 91% Accuracy 96% Accuracy
Badawi et al. [308]	2018		IMU	ADI	18	-	DT KNN MLP NB RF SVM	87% Accuracy 79% Accuracy 83% Accuracy 82% Accuracy 90% Accuracy 89% Accuracy
Scheurer et al. [315]	2017	Classification of firefighter movements	MIMU	firefighter exercises	11	LOSO	KNN XGB SVM	62% Accuracy 73% Accuracy 72% Accuracy
Zebin et al. [309]	2017		IMU	ADL	20	5-fold CV	DT Ensemble KNN LDA SVM	90% Accuracy 95% Accuracy 87% Accuracy 72% Accuracy 94% Accuracy
Ghazali et al. [326]	2018	Classification of sports activities	IMU		10	10-fold CV	DT Ensemble KNN LDA SVM	85% Accuracy 90% Accuracy 88% Accuracy 84% Accuracy 91% Accuracy
De Leonardis et al. [334]	2018		MIMU	ADL	15	5-fold CV	ANN DT KNN NB SVM	91% Accuracy 86% Accuracy 93% Accuracy 92% Accuracy 92% Accuracy

Std. Deviation: Standard Deviation, RateNorm: Rate of acceleration for normal running, RateBio: Rate of acceleration for running with biofeedback.

Table 5.6: Continued: Selection of studies examining the use of machine learning and inertial sensors/kinematic data for the prediction of activity recognition.

Author	Year	Study Objective	Relevant Data Type	Movement Captured	Number of Subjects	Validation Method	Classifier Used	Prediction Accuracy
Feng et al. [335]	2015		IMU	various sedentary, household, moderate, vigorous activities	17	10-fold CV	DT Ensemble KNN NB SVM	92% Accuracy 93% Accuracy 85% Accuracy 87% Accuracy 84% Accuracy
Nurhanim et al. [311]	2017		MIMU	ADL	30	10-fold CV	SVM	99% Accuracy
Tunca et al. [318]	2020	Fall Detection	IMU	Walking	76	LOSO CV	HMM LSTM MLP RF SVM	0.91 AUC 0.94 AUC 0.9 AUC 0.89 AUC 0.86 AUC
Özdemir and Barshan [319]	2014	Fall Detection	IMU	ADL	14	10-fold CV	ANN DWT KNN LSTM NB SVM	96% Accuracy 98% Accuracy >99% Accuracy >99% Accuracy >99% Accuracy >99% Accuracy
Delval et al. [320]	2020	Fall Detection	Kinematics	Walking	174	5-fold CV	SVM	97% AUC
Nahar et al. [321]	2020	Fall Detection	Accelerometer		6	10-fold CV	DT LogReg RF RepTree	97.4% Accuracy 96.5% Accuracy 95.6% Accuracy 96.3% Accuracy
Gillain et al. [322]	2019		IMU	Walking	96	10-fold	DT	84% Accuracy
Nahiduzzaman et al. [323]	2020		IMU		-	5-fold	RF RNN SVM	97% Accuracy 97% Accuracy 98% Accuracy
Ribeiro et al. [324]	2016	Detection of fainting and epileptic seizures	Accelerometer	Walking, standing, fainting and seizure	5	-	DT-1 DT-2 KNN NB SVM	99% Accuracy 99% Accuracy 99% Accuracy 85% Accuracy 88% Accuracy
de Pinho André et al. [336]	2019	Classification of correct movements	MIMU	ADL	12	LOSO	RF	87% Accuracy
Giggins et al. [328]	2014	Classification of correct movements	IMU	Rehabilitation exercises	58	LOSO	LogREg	94% Accuracy
Whelan et al. [327]	2016	Classification of correct movements performed	MIMU	Single leg squat	83	Repeated random sets	RF	78% Accuracy
Whelan et al. [327]	2017	Classification of correct movements performed	MIMU	barbell squat	55	LOSO	RF RF	61-66% Glob-Accuracy 79-83% pers-Accuracy
Chen et al. [329]	2013	Classification of correct movements performed	Accelerometer	Rehabilitation exercises	10	10-fold CV	DT	90% Accuracy

Std. Deviation: Standard Deviation, RateNorm: Rate of acceleration for normal running, RateBio: Rate of acceleration for running with biofeedback.

5.3.4 Prediction of Walking or Running Variables

A number of studies have examined the use of machine learning and inertial sensors or kinematics for the prediction of walking or running variables (not injury prediction). Twenty studies have been included in this review (see table 5.7) to provide an insight into the range of uses for machine learning with technique data (inertial and/or kinematic). Regression studies were also included as many kinematic variables (joint angles, joint loading, etc.) during walking or running are predicted or estimated as continuous variables. Studies have examined a variety of problems including prediction of muscle fatigue in walking [337, 338] and running [339–341], runner skill level classification [342, 343], classifying running conditions of running speed [344], terrain type classification during running [345] and the prediction of lower-body gait kinematics including estimation of joint angles [346–350], moments [348] and power [351] along with ground reaction forces [349] and maximal vertical loading rate [352]. Overall, 8 unique classifiers were used across the 20 studies, in order of most evaluated these classifiers were: Linear Regression, SVM, Neural Network, Deep Neural Network, Random forest, Ensemble methods, Naive Bayes, K-Nearest neighbour. Linear regression, SVM, and neural networks were the most used classifiers appearing 6 times each across these 20 studies.

Due to the mix of regression and classification tasks, there are a range of different measures used to evaluate classifier performance. Comparing standard accuracy across the classification studies the range of accuracies for each classifier were: SVM (66-98% Accuracy), Neural Network (97%), Deep Neural Network (96%), Random forest (75-100% Accuracy), Ensemble methods (97%), Naive Bayes (61%) and K-Nearest neighbour (62%).

Overall, across the 20 studies the level of prediction accuracy was high with the best classifier of each study (which reported standard accuracy for classification) performing approximately >=80%. These generally high levels of accuracy across prediction of various walking and running variables show that there are strong predictors within these specific tasks which can help to differentiate in the classification or regression problems.

In this section a selection of existing research was explored with respect to injury prediction, disease prediction, activity recognition, and prediction of a variety of running/walking applications. This section provides insight into the type of machine learning algorithms which have been heavily used with machine learning and technique data (inertial sensor and kinematic) and guides the selection of machine learning classifiers used in this thesis. This allows for a broad insight into the general high levels of accuracy which can be achieved using supervised learning, however, a clear observation

indirectly across studies shows that the performance of any given classifier is highly dependent on the problem, the extracted features and the specific study design followed. It can be seen from Table 5.1 that 7 of the studies examined predictions related to ACL injury, where machine learning models were trained to differentiate between healthy and ACL recovered limbs. Although each of these seven studies had the same general aim of classifying ACL injured and non-injured subjects, no single best classifier was found to excel at predicting ACL injury risk. It can be seen in table 5.1 that five unique classifiers were chosen as the best classifier across the six studies (or the only classifier examined in some cases). This is especially interesting comparing Mandalapu et al. [263], Richter et al. [267] and Tedesco et al. [268], all of which compared a number of different classifiers which overlap between studies. This variation in best classifier identified for the same classification task can be explained by differences in study design, for example, each study trained their model with data collected on different movements (walk and jog [263], jump and hop [267], run and sidestep [268]) while utilising different features/variables for prediction. This makes comparing classifier performance difficult across studies, as can be seen by [265] achieving 97% accuracy using an SVM classifier, while [263] reported SVM to perform the worst of 6 different machine learning models (46% accuracy) where both studies examined ACL injury risk. The No Free Lunch Theorem [353] refers to the idea that there does not exist a single superior optimisation algorithm (machine learning classifier) that can be blindly applied to all optimisation problems. This is why in many studies (including this thesis) a number of supervised learning algorithms were examined to gain insights into which algorithms were better suited to this specific problem and had better prediction performance.

Table 5.7: Selection of studies examining the use of machine learning and technique data for the prediction of walking or running variables.

Author	Year	Study Objective	Number of Subjects	Movement Captured	Relevant Data Type	Classifier Used	Validation Method	Problem Type	Prediction Accuracy
Op De Beéck et al. [339]	2018	Detection of Fatigue	29	Running	MIMU	ANN GradBoost LinReg-EN LinReg-LASSO	5-Fold CV	Regression	2.9 MAE 1.9 MAE 2.1 MAE 2.1 MAE
Marotta et al. [340]	2021	Detection of Fatigue	8	Running	MIMU	RF	LOSO CV	Classification	90% Accuracy
Guaitolini et al. [338]	2020	Detection of Fatigue	13	Walking	MIMU	SVM	LOSO	Classification	85% Accuracy
Zhang et al. [337]	2014	Detection of Fatigue	17	Walking	IMU	SVM-personal SVM-global	5-Fold CV	Classification	96% Accuracy 90% Accuracy
Buckley et al. [341]	2018	Detection of Fatigue	21	Running	MIMU	KNN-glob NB-glob RF-glob RF-pers SVM-glob	LOSO/10-fold CV	Classification	62% Accuracy 61% Accuracy 75% Accuracy 100% Accuracy 66% Accuracy
Gholami et al. [354]	2020	Prediction of joint angles	10	Running	Accelerometer	CNN	LOSO CV	Regression	3-7 degrees RMSE
Chow et al. [355]	2021	Prediction of joint angles	10	Running	IMU	CNN	LOSO/10-fold CV	Regression	3.6% to 10.8% RMSE
Zhang et al.[356]	2020	Predict gait parameters	14	Walking,Running	IMU	SVM SVM	10-fold CV	Regression	1-2 MAE (walking) 3-5% MAE (running)
Wouda et al. [350]	2018	Estimate vGRF and joint angles	8	Running	IMU	ANN ANN	-	regression	Knee angles RMSE <5o vGRF <0.3 RMSE degrees BW
Chaaban et al. [349]	2021	Predict vGRF and Knee Biomechanics	26	Jump Landing	IMU	LinReg	10-fold CV	Regression	4.6-10.2% RMSE
Lim et al. [347]	2019	Estimate lower limb Kinematics	7	Walking	IMU	ANN	LOSO	Regression	2-3 degrees RMSE
Mundt et al. [348]	2019	Prediction of joint angles and moments	75	Walking	IMU	FFNN FFNN LSTM LSTM	Held-out set	Regression	1.6 degrees nRMSE (angles) 15 m/s2 nRMSE (moments) 2.2 degrees nRMSE (angles) 12 m/s2 nRMSE (moments)
Jiang et al. [351]	2019	Estimate joint angle power	9	Walking	IMU	RF	LOSO	Regression	0.06 - 0.13 RMSE (w/kg)
Derie et al. [352]	2020	Estimation of VLR	93	Running	Accelerometer	XGB LinReg-EN LinReg-LASSO	LOSO	Regression	5.4 BWs-1 MAE 7.4 BWs-1 MAE 7.5 BWs-1 MAE
Strohrmann et al. [343]	2012	Classify running skill level	13	Running	IMU	SVM	LOSO	Regression	77% Accuracy
Liu et al. [342]	2020		30	Running	IMU	MLP	5-fold	Classification	97% Accuracy
Dixon et al. [345]	2019	Classification of outdoor terrain types	29	Running	Accelerometer	CNN XGB	-	Classification	96% Accuracy 97% Accuracy
De Brabandere et al. [357]	2020	Estimate Joint Loading	10	Walking, Stand-up Sit down, lunges, stairs	IMU	LinReg	LOSO	Regression	0.3-0.5% MAE
Stetter et al. [358]	2020	Estimate Joint moments	13	Walking, Running, Turning movement	IMU	ANN	LOSO	Regression	21-30% nRMSE
Benson et al. [344]	2018	Classification of running speed conditions	44	running	accelerometer	SVM	10-fold	Classification	97.5% Accuracy

ANN: Artificial Neural Network, LinReg: Linear Regression, MLP: Multi-Layer Perceptron, SVM: Support Vector Machine, XGB: Gradient Boosting (ensemble)

Chapter 6

Datasets: Data Collection

6.1 Introduction

This section outlines the collection of three different datasets which were used for the studies within this thesis. The datasets were used for examining the accuracy of orientation estimation, providing biofeedback to alter running technique, and determining predisposition to injury, as described in Chapters 7, 8 and 9, respectively.

Inertial Data

Inertial sensor data are central to the datasets used in each study of this thesis. These data are continuous time-series signals which are measured on axes X, Y and Z for each sensor within the sensing unit (e.g., accelerometer and gyroscope). Visual examples of these signals are included in figure 6.1.

6.2 Dataset 1: Determination of Inertial Sensor Orientation Accuracy

6.2.1 Overview of Experiment

Twenty participants were recruited for the study which involved capturing data on participants running on a treadmill for 30 minutes and walking for 2 minutes. The participants wore 3 inertial sensors, one placed on the sacrum, thigh and tibia. The inertial sensors recorded gyroscope and accelerometer data throughout the experiment, which were used to calculate segment orientation. The data recorded using the inertial sensors were compared to an optical motion analysis system, Vicon (Vicon Oxford Metrics, Oxford, United Kingdom), which is seen as a gold standard for accurately measuring human

(a) X-axis Accelerometer

(b) X-axis Gyroscope

(c) Y-axis Accelerometer

(d) Y-axis Gyroscope

(e) Z-axis Accelerometer

(f) Z-axis Gyroscope
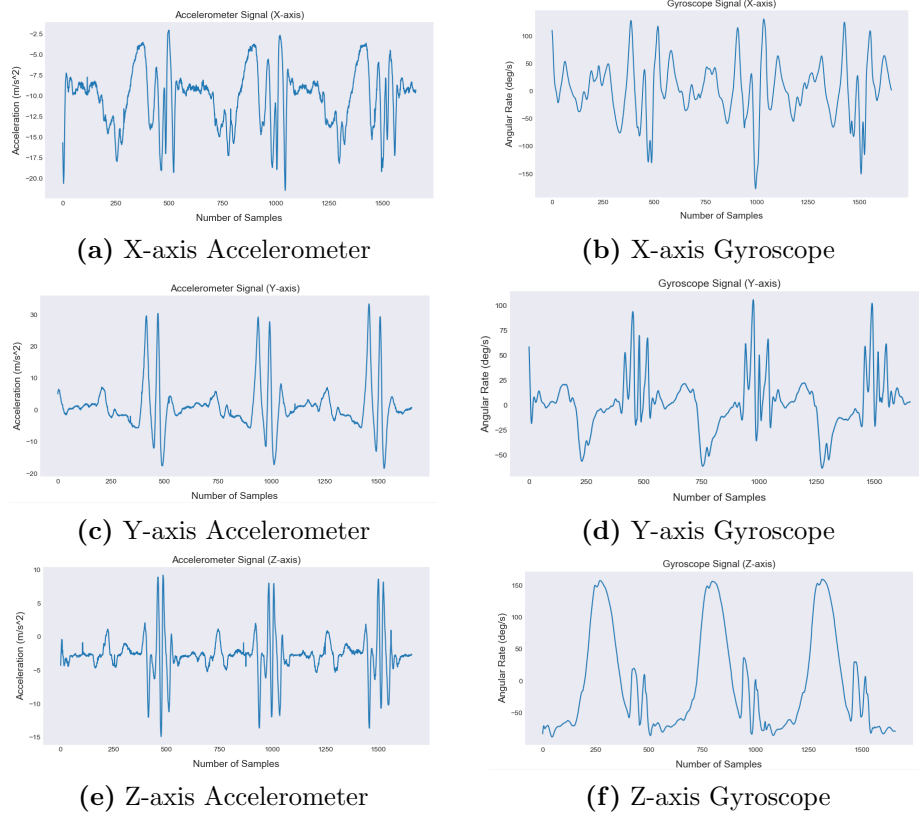
Figure 6.1: Continuous time-series accelerometer and gyroscope inertial data.

movement [359]. An additional inertial sensor (master sensor) was connected to the Vicon optical motion analysis system, to allow for synchronisation between the data of both systems.
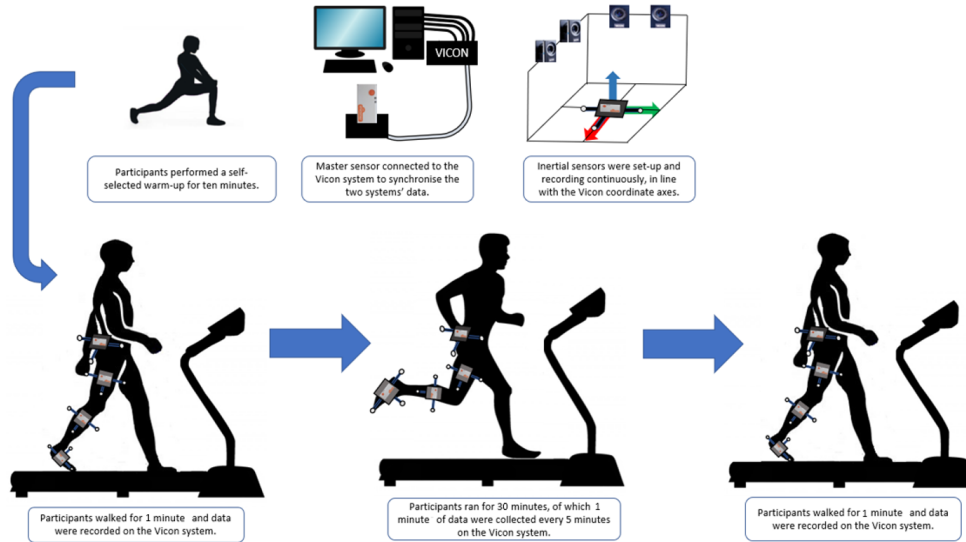
Figure 6.2: Dataset 1: Determination of Inertial Sensor Orientation Accuracy.

6.2.2 Equipment

The data in this experiment were captured using 3 wireless inertial sensors (Shimmer Sensing, Dublin, Ireland) and an optical motion analysis reference system (Vicon Oxford Metrics, Oxford, United Kingdom). The inertial sensors used were Shimmer3 inertial measurement units which contain a tri-axial accelerometer, tri-axial gyroscope and tri-axial magnetometer (see figure 6.3). The tri-axial magnetometer was not enabled during data collection. Inertial data and optical reference data were sampled at 442Hz as this was the highest sampling rate shared by both the inertial sensors and the optical reference system, which aided the synchronisation of both systems' data. The inertial sensors are 51mm x 34mm x 14mm in size (figure 6.4). Optical motion data were captured using 17 high speed cameras (Vicon Oxford Metrics, Oxford, United Kingdom) which track the position of retro-reflective markers attached to the participant. Segment orientation data calculated using the inertial sensors were compared to the orientation output from the optical motion analysis system.
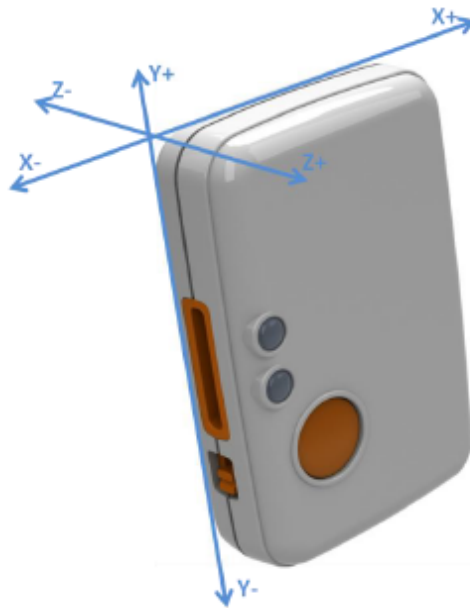
Figure 6.3: Axes of Shimmer3 inertial sensor.

6.2.3 Participants

Volunteers were recruited from undergraduate and postgraduate students and staff at Dublin City University on the Glasnevin Campus. The study was advertised by posters which were displayed at different locations around the DCU Glasnevin campus. The poster provided an overview of the study and contact details if anyone wished to find out more about the study. Twenty volunteers (12 male, 8

Figure 6.4: Shimmer3 IMU.

female) were recruited for the study with a mean (standard deviation) age, height and body mass of 29.3 (4.7) years, 1.76m (0.11) and 69.1 kg (13.6), respectively. Participants involved in the study were recreational runners between the ages of 18-40 years old, who had experience running on a treadmill. A recreational runner was defined as any runner who ran 10km per week or more, for more than 6 months [360]. Additionally, the participant must have had no lower-limb or back injuries in the past 6 months prior to the experiment. A single inertial sensor and a set of three optical retro-reflective markers (figure 6.5) were placed on each of the participant's analysed body segments on the right-side of the body; tibia, thigh, sacrum. Ethical approval was granted for the study by the Dublin City University Research Ethics Committee. Participants were given a Plain Language Statement explaining the tasks involved in the experiment, the data that would be collected and what the data would be used for. Each participant signed an Informed Consent form before being involved in the study.

6.2.4 Data Collection Protocol

The inertial sensors were calibrated using the Shimmer 9DOF Calibration Application (v2.10) and configured for the experiments using Shimmer Consensys Basic (v1.5.4). Prior to being placed on the participant, the inertial sensors were set-up to start recording while on the floor, aligned with the coordinate system axes of the Vicon optical motion capture space. This was to align the starting reference orientation frame of the inertial sensors with that of the Vicon optical motion analysis coordinate system, synchronising the starting 'zero' orientation for both the inertial sensors and optical motion capture system.

A master inertial sensor (not attached to the participant) received a signal from the Vicon system each time the Vicon system started and stopped data collection. This signal was used to synchronise

Figure 6.5: Plastic housing which was taped to each body segment. This was used to align the data collected by inertial sensor (within the box) and the retro-reflective markers (attached to the box).

the data between the two capture systems (Shimmer and Vicon).

The inertial sensors and Vicon retro-reflective markers were placed on the participant's body segments; tibia, thigh, and sacrum (see figure 6.2). The tibia sensor was placed midway between the knee and ankle joint, with the X-axis aligned with a line between the two joint axes and placed in the sagittal plane of the segment (X+ axis pointed downwards). The thigh sensor was placed midway between the hip and knee joint, with the X-axis aligned with a line between the two joint axes, placed in the sagittal plane of the segment (X+ axis pointed downwards). The sacrum sensor was placed midway between the posterior superior iliac spines, with the X-axis aligned with the spine and placed in the frontal plane of the segment (X+ axis pointed downwards). The inertial sensors were placed in a plastic housing that held the inertial sensors in place, and had optical markers attached to it at the same time (see figure 6.5). Each inertial sensor was centered in the box (at the origin of the local coordinate system formed by the trio of retro-reflective markers) to allow for alignment of sensor data and retro-reflective marker data. This set up was selected rather than placing the inertial sensors and retro-reflective markers directly on the skin, as it allowed for direct comparison between the orientation calculated by the inertial sensor data and the reference optical motion analysis system data (Vicon), without introducing errors associated with skin movement.
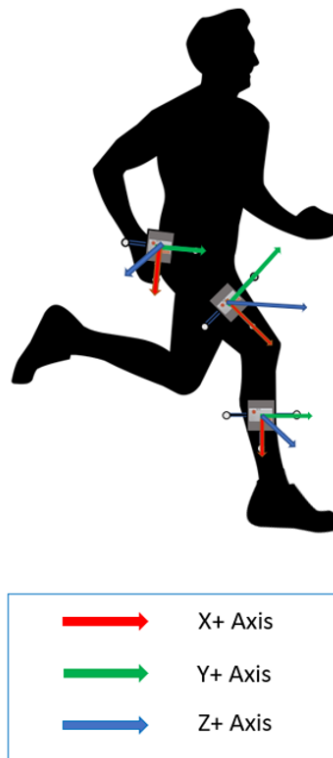
Figure 6.6: Sensor axes and sensor placement.

The inertial sensors recorded continuously from the initial 'zero' orientation (on the floor at origin) at the start of the experiment, until they were turned off at the end of the experiment. Vicon optical motion analysis data were recorded at specific intervals throughout the experiment. The orientation calculated in the recorded Vicon data intervals were later compared to the orientation calculated with inertial sensor data by synchronising the two systems' data.

Participants completed a self-selected warm-up including stretches for approximately ten minutes. They then walked on the treadmill at a self-selected pace, which was recorded for one minute. After one minute of walking, participants increased their speed to their self-selected running speed and ran continuously for 30 minutes. The running pace remained unchanged for the duration of the running phase of the experiment. Running data were recorded on the Vicon system every five minutes from the start of the run e.g., recordings at 0 minutes into run, 5 minutes...25 minutes, 30 minutes into run, with each recording lasting one minute. After 30 minutes, participants slowly reduced their speed to their original selected walking pace, and this final walk data was collected for 1 minute. The data from each inertial sensor were recorded to an internal SD card on board the device. Vicon system data and inertial sensor data were used to calculate segment orientation during the experiment for both walking and running.

6.3 Dataset 2: Real-time Biofeedback to Reduce Impact Loading during Running

6.3.1 Overview of Experiment

Sixteen participants were recruited for the study which involved capturing data on participants running in an outdoor environment. Participants were separated into two groups, Group A and Group B. Three inertial sensors were used on participants in both groups. Both groups (A and B) wore sensors on the sacrum and the tibia in order to determine the impact accelerations (impact loading assessment sensors). The third sensor provided real-time biofeedback to the participants. The biofeedback sensor streamed inertial data to a mobile phone app (SoftRun) via Bluetooth in real-time to provide the biofeedback to the runner. Group A had the biofeedback sensor placed on the tibia in order to provide impact acceleration-based feedback, and Group B had the biofeedback sensor placed on the thigh in order to provide (segment) orientation-based feedback. When the participant's data on the biofeedback sensor exceeded a set threshold, feedback was given in the form of an audio beep to inform them in real-time to 'run softer'. The impact load data collected during running with and without biofeedback was used to determine whether the two types of biofeedback helped the runner to alter their technique, and which of the two forms of biofeedback (impact acceleration based or segment orientation based) were more effective at reducing impact loading on the body.
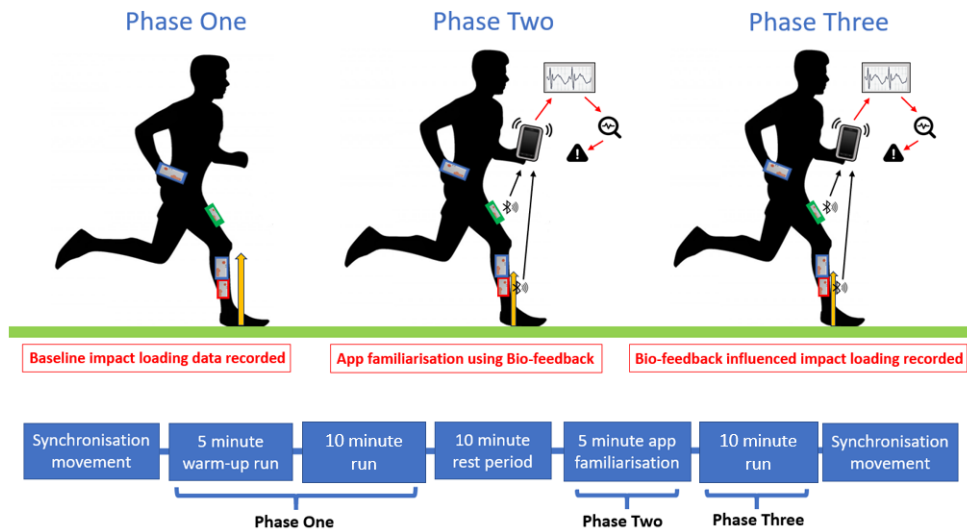
Figure 6.7: Dataset 2: Real-time Biofeedback to Reduce Impact Loading during Running

6.3.2   Equipment

The data in this experiment were captured using three inertial sensors (Shimmer Sensing, Dublin, Ireland). The tri-axial magnetometer was not enabled during data collection. The impact loading assessment sensors at the sacrum and tibia were sampled at 512Hz, and the biofeedback sensor was sampled at 100Hz for real-time streaming to the mobile phone. This experiment also required the use of a mobile phone with a specialised biofeedback app installed (SoftRun), which was developed by the research candidate for this study.

6.3.3   Participants

Volunteers were recruited from undergraduate and postgraduate students and staff at Dublin City University on the Glasnevin Campus. The study was advertised by posters which were displayed at different locations around the DCU Glasnevin campus. The poster provided an overview of the study and contact details if anyone wished to find out more about the study. Sixteen participants (equal numbers of male and female) were recruited for each group in the study with a mean (standard deviation) age, height and body mass of 26.1 (6.2) years, 1.73m (0.10) and 67.5 kg (11.4), respectively. Participants involved in the study were recreational runners between the ages of 18-40yrs old, who would run two or more times a week as part of a sport or fitness regime. A recreational runner was defined as any runner who ran 10km per week or more, for more than 6 months [360]. Additionally, the participant must have had no lower limb injuries in the past 6 months prior to the experiment. Participants were randomly assigned to one of the two study groups (A and B). Participants wore three inertial sensors which collected the data during the study. In both groups the impact loading assessment sensors were placed on the sacrum and the lateral aspect of the distal fibula. In Group A, the biofeedback sensor (third sensor) was placed on the medial distal aspect of the tibia. In Group B, the biofeedback sensor was placed on the anterior aspect of the distal end of the thigh. Ethical approval was granted for the study by the Dublin City University Research Ethics Committee. Participants were given a Plain Language Statement explaining the tasks involved in the experiment, the data that would be collected and what the data would be used for. Each participant signed an Informed Consent form before being involved in the study.
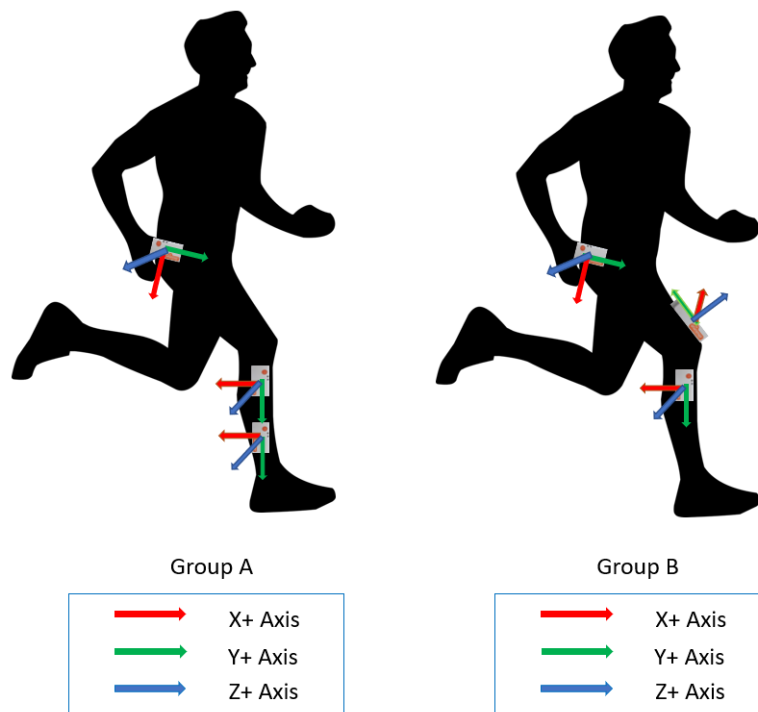
Figure 6.8: Sensor axes and sensor placement for groups A and B.

6.3.4 Data Collection Protocol

The inertial sensors were calibrated using the Shimmer 9DOF Calibration Application (v2.10) and configured for the experiments using Shimmer Consensys Basic (v1.5.4). Participants were separated into two groups, Group A and Group B. Both groups had sensors placed on the body to determine impact loading on the segments; sacrum and tibia (impact loading assessment sensors). A third sensor was placed on the tibia for Group A, and on the thigh for Group B, for the purpose of providing biofeedback (biofeedback sensor). A different form of biofeedback was provided to each group. The biofeedback sensor streamed inertial data to a mobile phone app in real-time to provide feedback based on impact accelerations to Group A (tibia biofeedback sensor), and to provide feedback based on segment orientation to Group B (thigh biofeedback sensor).

The data collection was split into three phases of running, with data collected only for the first and last phases, Phase One and Phase Three (figure 6.7). The first running phase (Phase One) captured normal baseline values from the participants' impact loading assessment sensors. The baseline biofeedback values were a measure of maximum peak impact acceleration of the tibia for Group A, and the thigh angle at initial contact for Group B. Biofeedback was not provided to the participants during Phase One. The baseline biofeedback values were collected in Phase One with the aim of reducing the magnitude of the measured values in Phase Three using the biofeedback provided. The maximum peak impact acceleration of the tibia was aimed to be reduced by 10% (Group A) and the thigh angle at initial contact was aimed to be reduced by 5 degrees using biofeedback. Phase Two allowed the participants to get used to using the mobile phone app and the biofeedback it provided while running. The data from this phase were not collected as part of the study. In Phase Three, the final running phase, inertial data from the participants' impact loading assessment sensors were captured while the participants were provided with biofeedback from the mobile phone app. The data from impact loading assessment sensors were recorded to an internal SD card, and the data from the biofeedback sensor was streamed in real-time to a mobile phone app over Bluetooth.

Before starting Phase One, the participants were given a demonstration of the 'SoftRun' app which would provide biofeedback for Phase Two and Three of this study. To begin Phase One, the impact loading assessment sensors were set to record inertial data and the biofeedback sensor was switched on and connected to the app. Biofeedback from the mobile phone app was not provided to the participant in this phase. To synchronise the data across inertial sensors, participants stood still for 10 seconds and performed 5 stiff jumps followed by standing still for another 10 second period.

This was repeated at the start and end of each phase of the data collection. Participants completed warm-up run for approximately 5 minutes. After this, participants ran for 10 minutes at a self-selected pace. Participants chose a pace that they would typically adopt for a normal 20-40 minute run. Five minutes into the self-selected pace run, 20 seconds of data were recorded from the biofeedback sensor by the researcher using the mobile phone app. This was collected in order to measure the maximum peak impact acceleration (Group A) or to measure the thigh angle at initial contact (Group B) in that time-frame. The measures collected in the 20 second window were considered representative of the whole data collected. After 5 minutes the runner stopped and rested for 10 minutes before the second running phase.

For the duration of Phase Two and Three the participant was in possession of the mobile phone with the 'SoftRun' app for biofeedback. Starting Phase Two, the participants were given the mobile phone and reminded of how to use the 'SoftRun' app. The participants were given 5 minutes to allow themselves to become familiar with the real-time biofeedback provided by the mobile phone app while running. A threshold was set in order to reduce the measured variables (10% reduction in maximum peak impact accelerations for Group A, and 5 degrees reduction in thigh angle at initial contact for Group B). If the participants' data on the biofeedback sensor exceeded the set threshold value, biofeedback was given in the form of an audio beep from the app in real-time which was an indication to 'run softer'.

Following this, participants started Phase Three by repeating a 10-minute run at a self-selected pace with the addition of biofeedback from the mobile phone. Participants were instructed to 'run softer' by following the advice to 'keep feet closer to the ground' and 'keep hips low to the ground' in order to reduce the number of feedback audio beeps received from the app. Acceleration peak and rate data from the loading assessment sensors were extracted across ten strides of running for Phases One and Three and compared, for each group A and B.

6.4 Dataset 3: Identifying the Risk of Injury and Injury Causative Factors using Machine Learning

6.4.1 Overview of Experiment

Three hundred and eleven (311) participants were recruited for the study which involved capturing data on participants running on a treadmill. The data were collected as part of a large running injury

study within the Running Injury Surveillance Centre (RISC 1), providing data for multiple PhD studies. The participants wore 7 inertial sensors, placed on the sacrum, and their left and right thigh, tibia and foot, which recorded accelerometer and gyroscope data. 32 retro-reflective markers were used to capture the runner's technique using Vicon (Vicon Oxford Metrics, Oxford, United Kingdom), an optical motion analysis system. The running data collected were recorded at a self-selected pace while participants were in an unfatigued state. Participants completed surveys detailing basic clinical data (sex, age, etc.) with information related to their running habits (running pace, distance) and injury history (previously injured, recently injured). Participants were tracked for running related injuries over 12-months, which established a dataset of prospectively injured and uninjured participants. The dataset outlined in this section was collected as part of a larger research study and was used in this thesis as an input to a machine learning model, with the aim to predict running related injuries from the prospective data of those who went on to become injured or remained injury free.
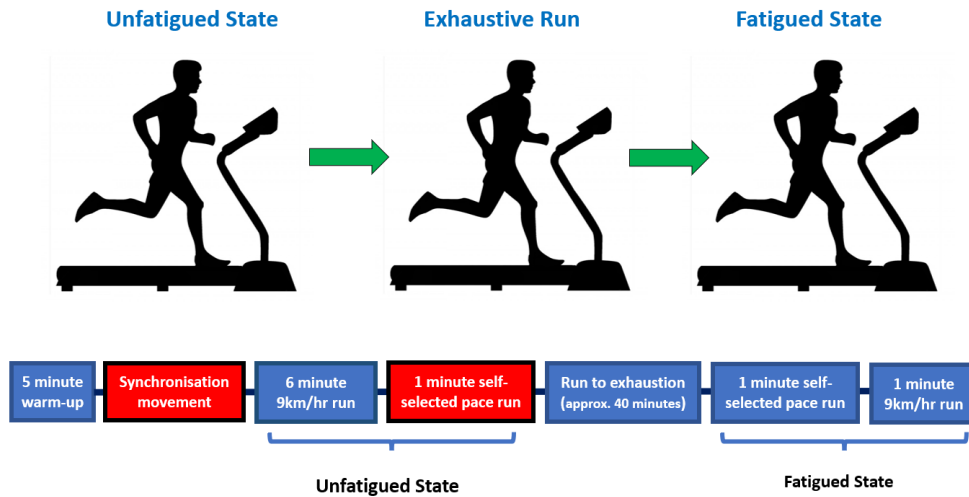
Figure 6.9: Dataset 3: Overall Protocol for Larger Data Capture. The portions of the data collection used in this study are highlighted in red.

6.4.2 Equipment

The data in this experiment were captured using 7 inertial sensors (Shimmer Sensing, Dublin, Ireland) and an optical motion analysis reference system (Vicon Oxford Metrics, Oxford, United Kingdom). Only accelerometer and gyroscope data were recorded. Inertial data were sampled at 512Hz and optical reference data were sampled at 200Hz. The seven inertial sensors were placed on the participant's sacrum, and their left and right thigh, tibia and foot. These were attached specifically to the following locations: dorsal aspect of the feet bilaterally, 2 cm above the talar dome, lateral aspect of the shank

bilaterally, 5 cm proximal to the lateral malleolus, lateral aspect of the thighs bilaterally, 10 cm proximal to the lateral knee joint and one IMU placed directly over the sacrum. Optical motion data were captured using 17 high speed cameras (Vicon Oxford Metrics, Oxford, United Kingdom) which track the position of 32 retro-reflective markers attached to the participant. The 32 retro-reflective markers were placed on the participant's lower limbs, trunk and pelvis following a custom Plug in Gait Vicon model composed of: ankle, heel, toe, tibia, knee, thigh, pelvis, posterior superior iliac spine, anterior superior iliac spine, shoulders, clavicle, sternum, C7 and C10 [361]. The Vicon Plug and Gait modelling routine used the retro-reflective markers to define rigid body segments (trunk, pelvis, thigh, tibia and foot), segment orientation and the joint angles between the segments. Before beginning motion capture, anthropometrical measurements were used as inputs to the Vicon Nexus software for bio-mechanical modelling (height, weight, leg length, knee width and ankle width).

(a) Front of body marker placement   **(b)** Back of body marker placement

Figure 6.10: Vicon retro-reflective marker placement on the participant.

6.4.3 Participants

Recreational and novice runners were recruited using social media, e-mails and posters. Three hundred and eleven participants were recruited as part of a larger research study examining running related injuries. Participants involved in the study were novice or recreational runners between the ages of 18-65 years old. The mean (standard deviation) age, height, body mass and body mass index for runners (95 male and 55 female) were 44.4 (8.5) years, 173.3 (9.2) cm, 72.3 (12.1) kg and 24.0 (2.8) kg/m2,

respectively. A recreational runner was defined as any runner who ran 10km per week or more, for more than 6 months [360]. A novice runner was defined as any runner who ran regularly for less than 6 months [362]. Participants were required to be injury free for three-months prior to the experiment and were required to not be involved in additional sports outside of running. This was to ensure that the injury data collected on the participants was primarily related to running based injury. Where multiple injuries occurred, only the first injury was recorded due as the following injuries maybe due to changing technique to compensate for the original injury. Ethical approval was granted for the study by the Dublin City University Research Ethics Committee. Participants were given a Plain Language Statement explaining the tasks involved in the experiment, the data that would be collected and what the data would be used for. Each participant signed an Informed Consent form before being involved in the study.

6.4.4 Data Collection Protocol

The inertial sensors were calibrated using the Shimmer 9DOF Calibration Application (v2.10) and configured for the experiments using Shimmer Consensys Basic (v1.5.4). The 32 Vicon retro-reflective markers were placed on the participants following a custom Vicon Plug in Gait model, which required marker placement on the participant, on their lower limbs, trunk and sacrum. The seven inertial sensors were placed on the participant, on their sacrum, and their left and right thigh, tibia and foot using tape. Detailed marker placement can be seen in figure 6.10 (a) and (b). Participants completed a self-selected warm-up including stretches for approximately 5 minutes. A stopwatch was used to determine the offset in the start of recording times between each inertial sensor once the sensors were set-up and recording. This was carried out for data synchronisation across sensors for later data analysis. The inertial sensors recorded accelerometer and gyroscope data for the duration of the data collection, while the Vicon motion analysis data were only collected to capture 1-minute of running at a self-selected pace. The Vicon data were used to extract joint angle data for the thorax, pelvis, (left and right) hip, knee, and ankle. Participants ran on the treadmill for one minute at a self-selected speed, during which data were collected using the Vicon motion analysis system. These data collected (see highlighted in red figure 6.9) were the only data used from this larger data collection for the study in this thesis.

The dataset collected were used to created 6 sub-datasets (inertial data only, kinematic data only [joint angles], clinical data only [age, sex, running pace, previous injury, etc.], and combinations of

inertial/clinical, kinematic/clinical and inertial/kinematic/clinical. These datasets were used to train seven machine learning classifiers to determine predisposition to injury based on the prospectively collected data. Inertial data measurements give an indication of body loading (accelerometer) and an indication of technique (accelerometer and gyroscope) during running, which are factors that are related to injury. The sub-datasets were used to investigate the benefit of using kinematic joint angle data (technique measure) versus inertial data (loading and quasi-technique measures) to determine predisposition to running related injuries.

6.4.5 Injury Tracking

Participants were asked to perform a once-off acute collection of their running data while running on a treadmill. Participants were tracked prospectively for 12-months and any participant who suffered a RRI in that time were assigned an 'injured' label in the dataset. A RRI was defined as *"any (training or competition) musculoskeletal pain in the lower limbs or lower back that causes a restriction/stoppage of running (distance, speed, duration, or training) for at least 7 days or 3 consecutive scheduled training sessions, or that requires the runner to consult a physician or other health professional"* [64]. Injury tracking involved contacting participants every 4-6 weeks to enquire about their current injury status. If an injury was reported, the participant was invited to have the injury inspected by two members of the research team who were qualified as a Chartered Physiotherapist and certified Athletic Trainer. If a participant could not attend this session, information regarding their injury complaints and any additional information given by their healthcare professional were recorded via phone call.

Chapter 7

Study 1: Determining and Improving the Accuracy of 6-Degrees of Freedom Orientation Estimation using Inertial Sensors during Running

7.1 Chapter Introduction

Running is extremely popular and has proven health benefits [4]; however, running injuries are extremely common with up to 80% of all recreational runners becoming injured each year [5]. Given the potential relationship between technique and injury [33], more insights into a runner's natural running technique would be beneficial to understand the factors leading to injury. Traditionally, human movement analysis has been carried out with highly accurate specialised equipment in laboratory-based environments. However, due to the high costs and lack of access to these clinical settings, few people benefit from the high accuracy provided by such systems. In addition, it has been shown that laboratory-based movement analysis does not necessarily reflect a runner's natural running technique [363–365]. It would be more valuable to be able to perform human movement tracking and analysis in an unconstrained way, in a runner's natural running environment. Inertial sensors provide a means of low cost, portable human movement tracking and have been used in a range of applications in the area of health, injury and sport.

The work in this thesis focused on a 6DOF approach of orientation estimation using inertial sensor data (accelerometer and gyroscope sensors) for calculating orientation and joint angles. The accelerometer data provides a measure of segmental loading on the attached body-part, and the combination of accelerometer and gyroscope data can be used for 6DOF orientation estimation. Additionally, 9DOF orientation is a popular approach where orientation can be calculated using the accelerometer, gyroscope and magnetometer data [191]. However, there are drawbacks to 9DOF orientation estimation where the inclusion of the magnetometer signal can cause orientation errors

caused by magnetic distortion in the local magnetic field due to soft and hard iron disturbances [6,7]. Although these errors can be reduced with calibration or compensation measures, it is not always possible for the magnetometer to be calibrated in the area of its intended use, especially if the runner is passing through areas with differing local magnetic fields [6,8]. Additionally, most of compensation techniques typically only allow for temporary disruption over short periods of magnetic distortion [9]. Therefore, the unpredictability of outdoor running in relation to local magnetic fields places doubt on the reliability of 9-DOF estimation, especially where signal complexity is analysed. A possible solution is the use of 6-DOF estimation, which ignores the magnetometer signal.

With the use of 6-DOF to calculate segment orientation and joint angles from inertial sensors, a number of challenges exist. There are three main challenges: signal drift [12], movement type (high impacts and speed) [8] and data loss [13]. Signal drift refers to the errors which accumulate in gyroscope-based orientation estimation and cause the signal to diverge from the true values over time. This occurs when sensor data with noise and bias errors (naturally existing within sensor data) are used to calculate segment orientation or joint angle estimates, and the step of integration causes these errors to become larger, greatly affecting the orientation estimate. Without compensating for signal drift, the amount of time that orientation can be accurately recorded is relatively short [167]. In order to provide useful insights, the sensor-based orientation estimates must be sufficiently accurate for lengths of time consistent with a sports training session (>30 minutes). The majority of 6DOF studies (as highlighted in the review section 3.1) perform short testing durations only. There is a need to understand the level of errors that exist over a longer duration of time. Therefore, the dataset used for studies in this Chapter (Dataset 1: Section 6.2) involved running data collected over a duration of 30 minutes, for the comparison of the effect of time on the accuracy of the approach taken.

Additionally, the type of movement being recorded by the sensors is important as it has been directly and indirectly shown that faster movements produce larger orientation errors [8, 197]. High impact and fast movement actions e.g., running, have higher amounts of force and vibration on the body/attached sensor, adding noise to the sensor readings [366]. These large impact accelerations can lead to distortion of the direction of the gravity vector (measured via the accelerometer) which is an important aspect of calculating 6DOF orientation and therefore leads to orientation errors [10, 11]. However, the magnitude of these errors (due to impact accelerations during running), and whether they can be reduced to a clinically acceptable level of error (less than 5 degrees [367, 368]), remains unclear. For useful insights the sensor data is required to remain sufficiently accurate during high

impact, fast actions as well as low impact, slow actions. Few studies involving inertial sensor data have directly examined estimating 6DOF segment orientation during running, with the majority of research investigating other aspects of running (e.g., estimation of running speed [17] or running gait symmetry [18]). For this reason, our studies have involved the collection of both running and walking data to directly examine the effect of impact presence and movement speed on the accuracy of the algorithm.

Data loss is a common issue affecting all sensors, including the inertial sensors used in this thesis. Data samples to be recorded can be missed by the internal processor of a sensor, for example when the sampling rate or the number of sensors enabled are high. To calculate orientation, the previous known orientation is used to attain the current orientation estimate, and so inaccuracies that exist may have a cumulative effect on subsequent orientation estimates [13]. The Madgwick filter [14] was used for 6DOF orientation calculation in this thesis. It is unclear how the Madgwick filter deals with data loss, if at all, as it does not appear to be taken into account in the algorithm. The effect of data loss on the algorithm accuracy is investigated in the experiments of Study 1.2 below (Section 7.4).

There may be ways to improve the 6DOF running orientation estimation to acceptable levels of accuracy. A number of studies using both 6DOF [7,210,216] and 9DOF [12,191,366] have examined the optimisation of the beta parameter within the Madgwick filter itself. This parameter has an important role in the reduction of signal drift error within the algorithm. This parameter represents the weight given to either the gyroscope or accelerometer data, where a beta value of zero would calculate orientation using the gyroscope only, and a beta value >0 represents the magnitude of influence from the accelerometer data. In essence, this parameter manages the success of the convergence of the algorithm during movements. An examination of the difference in running orientation provided by tuning this parameter was carried out in Study 1.3A (see section 7.5). Additionally, there may be another option to improve running orientation by changing the inner structure of the algorithm itself. The Madgwick filter relies on the use of Stochastic Gradient Descent to perform an optimisation scheme which aims to use accelerometer data to reduce the error in the gyroscope estimate. Other Stochastic Gradient Descent optimisation schemes exist (Momentum, RMSprop, Adam [369]) which are commonly used in the domain of artificial intelligence and deep learning. It is possible that using a different stochastic gradient descent optimiser e.g., momentum, may improve the running orientation estimates; no studies have investigated this to-date. This was explored in Study 1.3B (Section 7.6) of this thesis.

This chapter is divided into three main sets of experiments. Study 1.1 details the implementation and testing of the baseline complementary filtering algorithm [14] and compares this to ground-truth reference data to evaluate the level accuracy of the inertial sensor orientation against the reference system for walking and running over a 30-minute data collection session. Additionally, this chapter aimed to investigate approaches to improve the baseline orientation algorithm for better segment orientation accuracy, as examined in Study 1.2 and Study 1.3. Study 1.2 details the issue of data loss and how this could be dealt with within a data pre-processing pipeline. Study 1.3 (A and B) details the implementation of modified versions of the orientation algorithm used in Study 1.1 and assesses the difference in accuracy of these approaches against the baseline results using the ground-truth data.

7.2 Initial Common Methodology

The following initial methodology is common to each study (1, 2, 3A and 3B) described below.

7.2.1 Overview of the Data Collection

Inertial sensors were used to record tri-axial accelerometer and tri-axial gyroscope data on three body segments (sacrum, thigh and tibia). These data were input to an orientation algorithm to calculate human segment orientation. The Vicon system tracked reflective markers that were placed on the participant by using high speed cameras which were positioned around the perimeter of the capture space. Marker position data was extracted which provided data describing each marker as a three-dimensional (3D) point in the global Vicon system capture space [see figure 7.1 (a)]. From this marker data, a local coordinate system was created at each segment [see blue axes on figure 7.1 (b)], and the orientation of each of these local coordinate systems were used to track segment orientation. The data from the Vicon system was synchronised with the inertial sensor data so that the same recorded sections of data could be compared. Segment orientation was calculated for walking (1-minute at the beginning), running (1-minute recorded every 5-minute intervals for duration of 30 minutes), and for a final 1 minute of walking. Section 6.2 describes in detail the common dataset which was used for the experiments in this Chapter. Each of the studies below (1, 2, 3A and 3B) used the same experimental dataset so that the results of these sections could be directly compared.
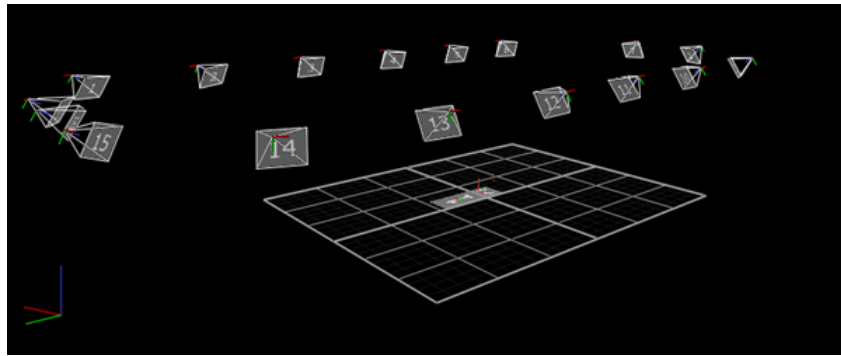
(a) Vicon system tracking capture space.
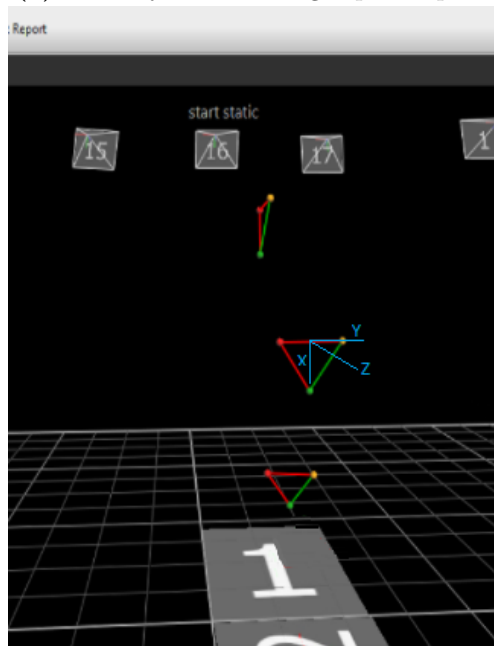
(b) Retro-reflective markers used for segment tracking.

Figure 7.1: Figure (a) shows the user interface from the Vicon tracking System showing the positioning of 17 high speed camera around the capture space. Figure (b) shows the triads of retro-reflective markers placed on each segment as identified via the Vicon tracking system. The blue axes (X, Y, Z) represent the local co-ordinate system which was created using the retro-reflective marker data.

7.2.2 The Madgwick Filter

As can be seen in the conclusion of the review in Section 3.6, there is a need for further research to assess the accuracy of orientation estimation during running, captured over an extended period of time. It appears that the Complementary filter is the most appropriate orientation estimation approach for this PhD thesis, as it is less computationally heavy for real-time applications when compared to other orientation estimation algorithms and has comparable levels of accuracy [162, 197, 199].

The Madgwick filter [14], which has been used in many studies estimating human movement [8, 191, 192], was implemented as the complementary filter of choice for this PhD thesis. The Madgwick

filter is designed to first calculate an orientation estimate using the gyroscope signal alone through integration, Qw.

$$^S\boldsymbol{\omega} = [0, \omega_x, \omega_y, \omega_z] \tag{7.1}$$

$$_E^S \dot{\boldsymbol{q}} = \frac{1}{2} \, _E^S \hat{\boldsymbol{q}} \otimes^S \boldsymbol{\omega} \tag{7.2}$$

$$_E^S \dot{\boldsymbol{q}}_{\omega,t} = \frac{1}{2} \, _E^S \hat{\boldsymbol{q}}_{est,t-1} \otimes^S \boldsymbol{\omega}_t \tag{7.3}$$

$$_E^S \boldsymbol{q}_{\omega,t} = {_E^S}\hat{\boldsymbol{q}}_{est,t-1} + {_E^S}\dot{\boldsymbol{q}}_{\omega,t}\Delta t \tag{7.4}$$

As gyroscope integration cannot exclusively be used to calculate orientation due to accumulating signal drift (see Section 3.2), the accelerometer signal is used to offset the error in the gyroscope based orientation estimate. The accelerometer signal measures both the direction and magnitude of gravity (according to the sensor frame) combined with the linear accelerations associated with the movement of the sensor. This means that in quasi-static conditions the magnitude and direction of the accelerometer signal measurements are mostly related to the field of gravity alone. Qa is the orientation estimate which aligns the known measured direction of accelerometer sensor measurements, with a pre-defined reference direction in the earth frame. This becomes an optimisation problem defined by minimising the difference between Ed and Ss (Eq. 6.5)

$$\min_{_E^S\dot{\boldsymbol{q}}\in R^4} \boldsymbol{f}(_E^S\hat{\boldsymbol{q}}, ^E\hat{\boldsymbol{d}}, ^S\hat{\boldsymbol{s}}) \tag{7.5}$$

$$\boldsymbol{f}\left(_E^S\hat{\boldsymbol{q}}, ^E\hat{\boldsymbol{d}}, ^S\hat{\boldsymbol{s}}\right) = {_E^S}\hat{\boldsymbol{q}}^* \otimes {^E}\hat{\boldsymbol{d}} \otimes_E^S \hat{\boldsymbol{q}} - {^S}\hat{\boldsymbol{s}} \tag{7.6}$$

$$_E^S\boldsymbol{q}_{k+1} = {_E^S}\hat{\boldsymbol{q}}_k - \mu \frac{\nabla\boldsymbol{f}\left(_E^S\hat{\boldsymbol{q}}_k, ^E\hat{\boldsymbol{d}}, ^S\hat{\boldsymbol{s}}\right)}{\left\|\nabla\boldsymbol{f}\left(_E^S\hat{\boldsymbol{q}}_k, ^E\hat{\boldsymbol{d}}, ^S\hat{\boldsymbol{s}}\right)\right\|}, k = 0, 1, 2 \dots n \tag{7.7}$$

$$\nabla\boldsymbol{f}\left(_E^S\hat{\boldsymbol{q}}_k, ^E\hat{\boldsymbol{d}}, ^S\hat{\boldsymbol{s}}\right) = \boldsymbol{J}^T\left(_E^S\hat{\boldsymbol{q}}_k, ^E\hat{\boldsymbol{d}}\right) \boldsymbol{f}\left(_E^S\hat{\boldsymbol{q}}_k, ^E\hat{\boldsymbol{d}}, ^S\hat{\boldsymbol{s}}\right) \tag{7.8}$$

$$\boldsymbol{f}_g \left({}_E^S \hat{\boldsymbol{q}}, {}^S \hat{\boldsymbol{a}} \right) = \begin{bmatrix} 2\left(q_2 q_4 - q_1 q_3\right) - a_x \\ 2\left(q_1 q_2 + q_3 q_4\right) - a_y \\ 2\left(\frac{1}{2} - q_2^2 - q_3^2\right) - a_z \end{bmatrix} \tag{7.9}$$

$$\boldsymbol{J}_g \left({}_E^S \hat{\boldsymbol{q}} \right) = \begin{bmatrix} -2q_3 & 2q_4 & -2q_1 & 2q_2 \\ 2q_2 & 2q_1 & 2q_4 & 2q_3 \\ 0 & -4q_2 & -4q_3 & 0 \end{bmatrix} \tag{7.10}$$

$$ {}_E^S \boldsymbol{q}_{\nabla,t} = {}_E^S \hat{\boldsymbol{q}}_{est,t-1} - \mu_t \frac{\nabla \boldsymbol{f}}{\|\nabla \boldsymbol{f}\|} \tag{7.11}$$

Madgwick uses Stochastic Gradient Descent to minimise the objective function **f**. Unlike in traditional Stochastic Gradient Descent where one sample (sensor measurements at time t) is iterated over a number of times to traverse the error surface for each measurement, in Madgwick's filter there is only one iteration of Stochastic Gradient Descent per sample at time t, in order to reduce the computational load of the calculations. Stochastic Gradient Descent is used to calculate a second estimate of orientation based on accelerometer alone by finding the orientation which minimises the difference between the measured and reference accelerometer vector. The algorithm fuses these two estimates of orientation to produce a more reliable orientation estimate, which highlights the benefits of sensor fusion. Through math-based assumptions the equations are reduced down to the fusion equation described in equations 7.11 - 7.13 (the derivations of which can be found in the original paper by Madgwick et al. [14]. This fused approach allows the accelerometer data to minimise the signal drift during low frequency movements when the gravity vector can be easily read from the data.

$$ {}_E^S \boldsymbol{q}_{est,t} = {}_E^S \hat{\boldsymbol{q}}_{est,t-1} + {}_E^S \dot{\boldsymbol{q}}_{est,t} \Delta t \tag{7.12}$$

$$ {}_E^S \dot{\boldsymbol{q}}_{est,t} = {}_E^S \dot{\boldsymbol{q}}_{\omega,t} - \beta {}_E^S \dot{\hat{\boldsymbol{q}}}_{\epsilon,t} \tag{7.13}$$

$$ {}_E^S \dot{\hat{\boldsymbol{q}}}_{\epsilon,t} = \frac{\nabla \boldsymbol{f}}{\|\nabla \boldsymbol{f}\|} \tag{7.14}$$

Without using an additional reference field direction (e.g., magnetic field using 9-DOF) to align

the final axis of rotation, a unique solution cannot be found, as there are infinitely many solutions on the yaw axis (Z-axis). While Madgwick's 6-DOF approach will only estimate drift reduced roll (X-axis) and pitch (Y-axis) angles, there are approaches which can be used to estimate yaw (Z-axis) angles.

7.2.3 Data Synchronisation

Inertial sensor data were downloaded from the sensor SD cards using Consensys software v1.4 provided by the inertial sensor manufacturer (Shimmer Sensing, Dublin). The inertial sensors were synchronised with each other by using a master-slave configuration, which required an additional inertial sensor (master) not attached to the participant. The Consensys software performed the data synchronisation across all sensors.

Synchronising the inertial sensor data with the Vicon system data was performed using the same master sensor (mentioned above) which received an input signal from the Vicon system during data collection each time recording started or stopped. The signal received by the master inertial sensor was similar to a square-wave signal accompanied by the timestamp at which each sample was received.

7.2.4 Processing of Inertial Sensor Data

For each participant, there were three separate inertial sensor data files for the experiment (sacrum, thigh and tibia) which were synchronised with the master sensor as described above. The master sensor was subsequently used to identify periods of data capture from the Vicon system for each segment data. The master data was only used to identify the recording section from the Vicon system, and time-synchronise the inertial sensors to each other.

Orientation was calculated for each segment data, which started from its 'zero' orientation until the end of the experiment. Orientation was calculated using Madgwick's 6DOF orientation algorithm [14]. The orientation algorithm used the gyroscope and accelerometer data as inputs (sampled at 442Hz). The fixed β parameter and starting orientation were initialised to $\beta = 0.033$ and quaternion = [1,0,0,0], respectively, as suggested for 6DOF in Madgwick's paper [14]. Orientation was represented in quaternions throughout all inertial data calculations as they are more computationally efficient in comparison to Euler angles and avoid singularities which can occur e.g., Gimbal Lock [177]. Once the entire segment's inertial data orientation was calculated, for each segment, the recording sections matching periods of data capture from the Vicon data were extracted for direct comparison between

the two systems (inertial sensor and Vicon). Data were only transformed from quaternions to Euler angles and rotation matrices once all orientations were calculated for numerical and visual analysis.

7.2.5 Processing of Vicon System Data

The Vicon data recordings were processed using the Vicon system software, Vicon Nexus v2.18. The data were processed to determine the 3D coordinates of each reflective marker on the participant. The participant data were screened to exclude any anomalies in the data. The screened marker data were filtered with a 4th order, low-pass Butterworth filter, with a cut off frequency of 15Hz. From this screened marker data, three local coordinate axes were created using sets of three 3D coordinates defining a segment (3 reflective markers) and vector mathematics [see blue axes in figure 7.1 (b)]. These coordinate axes define a local coordinate system within the global coordinate space. A local coordinate system was created at each segment and the orientation of each of these local coordinate systems were used to track segment orientation. Orientation data were initially represented as rotation matrices; however the data were converted to Euler angles for analysis. This was carried out on each segment for all participants.

7.3 Study 1.1: Comparison of the Inertial Sensors vs Optical Motion Analysis System

The aims study 1.1 was to initially determine the level of accuracy of Madgwick's Complementary filter [14] in calculating segment orientation for walking and running over an approximately 30-minute period while implementing a novel signal drift compensation mechanism. A gold standard optical motion analysis system, Vicon (Vicon Oxford Metrics, Oxford, United Kingdom), was used as reference data on which to calculate orientation error. To address the aim, the experiments outlined investigated the effect of movement type (walking and running) and duration of time (capture time) on the accuracy of the orientation estimates, and additionally the effect of (capture) time on estimated running orientation accuracy.

7.3.1 Methodology

7.3.1.1 Data Processing

The Madgwick 6DOF filter as described above was used to calculate segment orientation from the experiment using the inertial sensor motion captured gyroscope and accelerometer signals. Orientation estimates were computed in quaternion form for each segment of a participant (tibia, thigh and sacrum) and segmented into 1-minute sections (time-points synchronised with the reference data). The quaternion estimates in these segmented sections were converted to Euler angles (XYZ) for reference data comparison and visualisation purposes. These comparison sections were based on the timestamps of Vicon reference recordings, (i.e., start walk, run at 0 minutes, run at 5 minutes, ..., run at 30 minutes, end walk). This segmentation was carried out using the time-synchronisation of the datasets.

Due to the fact that Madgwick's 6-DOF algorithm does not account for drift in the yaw axis [14], a number of simple drift removal approaches were examined: linear de-trending, mean de-trending, high-pass filter de-trending and a number of polynomial curve de-trending techniques (quadratic, cubic, quartic and quintic). These techniques were examined on a subset of the dataset (ten segments chosen from ten participant's data. As the Madgwick 6DOF filter already handles drift removal on the X and Y axes (roll and pitch angles respectively), this analysis was only concerned with Z-axis (yaw angle) de-trending performance. Within each segmentation section the Z-axis sensor orientation data was fit with a de-trending technique (listed above), which centered the de-trended Euler angles around zero degrees (data centred around X-axis by de-trending, see figure 7.3). With the data centred around zero, a positive offset was applied to the data based on the Z-axis Vicon (reference) calculated Euler angle, in order to roughly align both estimates and calculate orientation error. This positive offset was obtained by calculating the RMSE between the Z-axis Vicon orientation data and the zero centred (de-trended) sensor Z-axis orientation data. In this case the Vicon acts as the additional reference to align absolute orientation. In a laboratory free environment this offset could be calculated before exercise using the average Z-axis Euler orientation angle collected in a few seconds when the person is standing still with the sensor attached to their leg. As can be seen in figure 7.2, the linear de-trending performed the best drift removal on the Z-axis and therefore was applied to the full dataset.
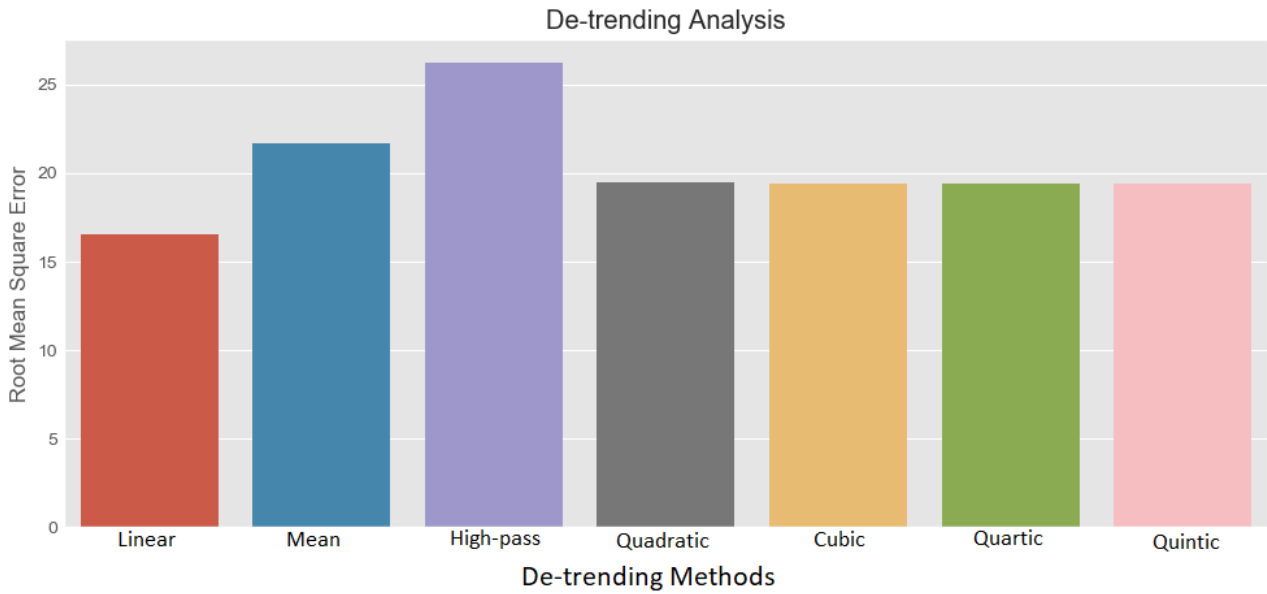
Figure 7.2: Analysis of a number of de-trending techniques were carried out on a subset of that dataset. It can be seen that linear de-trending had the lowest RMSE.

The inertial data used to calculate orientation (gyroscope and accelerometer) suffered from data loss throughout the recording. It seems that the Madgwick filter does not consider missing data samples in its calculation, though importantly calculating orientation relies on the last known orientation to be accurate. In order to keep the sensor data and Vicon data time-aligned in these comparison sections, gap filling was performed on the Euler angles to fill any missing data. This is what we will refer to as *post-filling* the data (post refers to gap-filling **after** after orientation is estimated). This post-filling was carried out using a linear interpolation. This was chosen as it was a simple approach for gap filling and has been used previously for data re-sampling in inertial sensor data [370, 371]. It is worth noting, that no studies appear to have mentioned how they account for data loss when aligning their calculated orientation with their reference orientation values. A Python implementation of Madgwick's 6DOF Filter was used by adapting the C code which can be found in Madgwick's report [372].
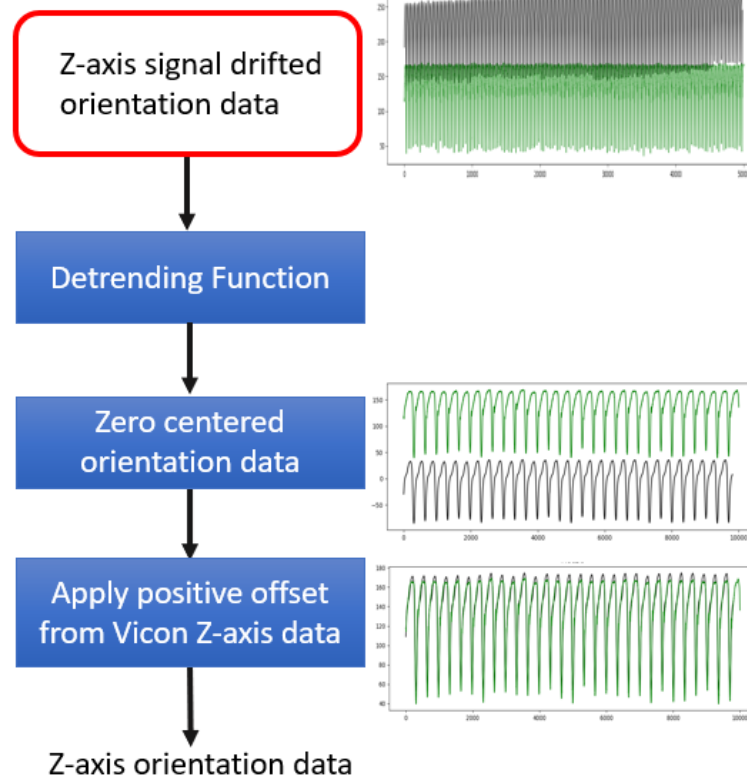
Figure 7.3: Z-axis de-trending flowchart of steps involved.

7.3.1.2 Data Analysis

RMSE was used to compare the inertial sensor orientation estimates with that of the reference data for all subject data. Data were analysed at a segment level by averaging across the X, Y and Z RMSE for each particular comparison section (i.e., walk start), this was performed due to the number of outliers that occurred in the data that could not be removed for analysis. For example, statistical analysis to assess the effect of capture time across time-points (run0, run10, run15, run20, run25, run30) could not be accurately performed if the error of a certain time-point was inflated for an individual participant due to outliers. Therefore, data were averaged across axes for an individual participant in order to remove as few participants as possible from the statistical analysis.

The distribution of the data was assessed by Shapiro-Wilk's test for normality, and it was found that the data was primarily non-normally distributed ($p < 0.05$). To examine the effect of movement type and capture duration on the orientation estimates, a non-parametric Friedman Test was used to evaluate the differences between (Walkstart, Run0, Run30, WalkEnd). Where a significant result was found, Wilcoxon Signed Rank tests were carried out between all paired combinations of (Walkstart,

Run0, Run30, WalkEnd) to discover where the significant difference existed. Another Friedman test was carried out to assess the effect of capture time on the accuracy of the running orientation data. This was examined to show whether the orientation accuracy changed as time progressed, especially in the Z-axis due to signal drift. A non-significant result of this test implies that the algorithm and Z-axis de-trending consistently manages signal drift effectively across time. Where Mauchly's test of Sphericity was violated, Greenhouse-Geisser correction of the degrees of freedom was used. The statistical analysis was completed using IBM SPSS Statistics 24. The level of significance was set at $p < 0.05$.

7.3.2 Results

An example of the Z-axis de-drifting protocol can be seen in figure 7.4 (a) - (c), where the inertial orientation data is successfully aligned with the reference. The magnitude of the average tibia orientation RMSE for roll (X), pitch (Y) and yaw (Z) were 26.2, 10.5 and 15.2 degrees, respectively. The magnitude of the average thigh orientation RMSE for X, Y and Z were 19.2, 4.8 and 11.8 degrees, respectively. The magnitude of the average sacrum orientation RMSE for X, Y and Z, were 22.7, 7.0 and 17.2 degrees respectively (see table 7.1). The results were consistent across segments, following a similar overall pattern of significance. A Friedman test examined the effect of Time (0 minutes, 30 minutes) and Movement type (walk, run) on the orientation accuracy via the conditions (Walkstart, Run0, Run30, WalkEnd). Each segment Friedmann test found a significant relationship among the variables and post-hoc analyses were carried out using Wilcoxon Signed Rank tests. Additionally, the second Friedman test examining the effect of Time on the running orientation accuracy, found that there was no statistically significant difference between measurements across time points, for all segments. P values and test statistics relating to the analyses can be found in tables 7.2 - 7.8.

RMSE (Degrees)	X	Y	Z
Tibia	26.2	10.5	15.2
Thigh	19.2	4.8	11.8
Sacrum	22.7	7.00	17.2

Table 7.1: Average segment running orientation RMSE (Degrees) accuracy across all axes.
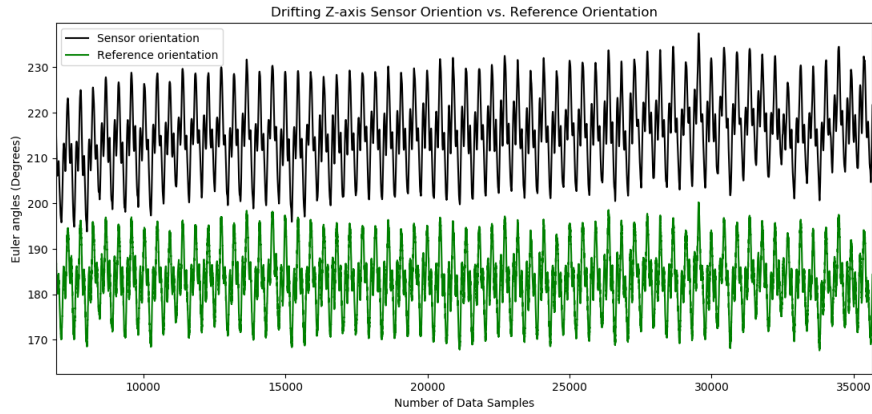
(a) Drifting Z-axis data
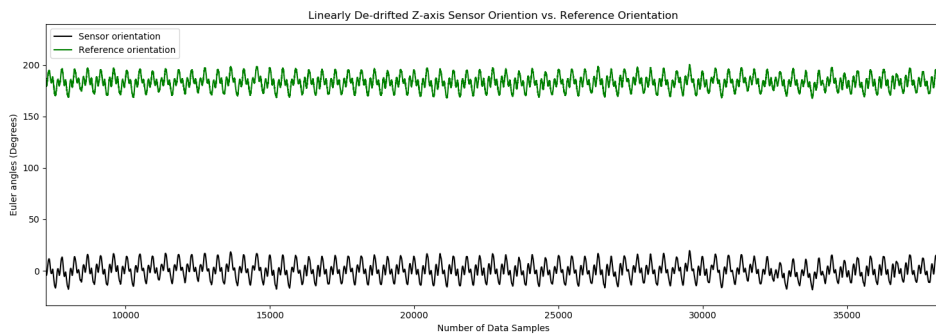
(b) De-drifted Z-axis data
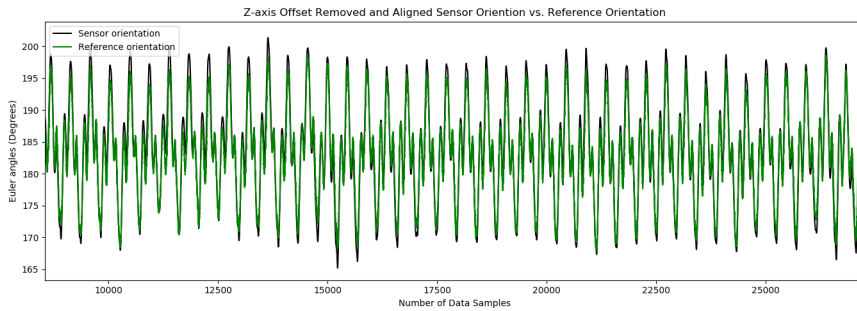
(c) Aligned Z-axis data

Figure 7.4: The pipeline process of removing the linear drift from the Z-axis resulting in de-drifted data. This de-drifted data is centered around zero, and so a positive offset (calculated using Vicon data) is applied in order to arrive at aligned signals.

Statistical Analysis Results of the Tibia Segment data

Results from the Friedman test showed that there was a statistically significant relationship ($\chi^2(20)$ = 15.1, p = 0.002) among the variables [Walkstart, Run0, Run30, WalkEnd]. Post-hoc analyses were carried out using Wilcoxon Signed Rank tests to examine every combination pair of the variables (Walkstart, Run0, Run30, WalkEnd). Each of the two pairs containing the same movement type were not significantly different [WalkStart - WalkEnd: (Z = -0.78, p = 0.43), Run0 - Run30: (Z = -0.64, p

= 0.53)]. The 4 remaining pairs of differing movement type were all statistically significant (p<0.05) [see table 7.2 for test statistics]. This suggests that the Movement type had a significant effect on orientation estimation where the faster movement, running, had higher errors. The full set of ranks and test statistics for this Friedman test can be found in tables 7.7 and 7.8.

The additional Friedman test, to evaluate whether there was a significant change in orientation accuracy over time, showed that there was no statistically significant relationship between measurements across time points (Run0, Run10, Run15, Run20, Run25, Run30) [$\chi^2(20)$ =1.86, p = 0.93]. The full set of ranks and test statistics for this Friedman test can be found in tables 7.5 and 7.6.

Table 7.2: Wilcoxon Signed Rank Test results for the tibia in analysing the statistically significant relationship within Time and Movement Type.

Test Pair	Z_{Tibia}	p_{Tibia}
WalkStart-Run0	-2.89	0.004*
WalkStart-Run30	-2.17	0.03*
WalkStart-WalkEnd	-0.78	0.43
Run0-Run30	-0.64	0.53
Run0-WalkEnd	-3.40	0.001*
Run30-WalkEnd	-2.99	0.003*

Where p: significance p-value, Z: Wilcoxon Signed Rank test statistic, * denotes statistical significance.

Statistical Analysis Results of the Thigh Segment data

Results from the Friedman test showed that there was a statistically significant relationship ($\chi^2(17)$ = 14.5, p = 0.002) among the variables [Walkstart, Run0, Run30, WalkEnd]. Post-hoc analyses were carried out using Wilcoxon Signed Rank tests to examine every combination pair of the variables (Walkstart, Run0, Run30, WalkEnd). Each of the two pairs containing the same movement type were not significantly different [WalkStart - WalkEnd: (Z = -1.07, p = 0.29), Run0 - Run30: (Z = -1.63, p = 0.10)]. The 4 remaining pairs of differing movement type were all statistically significant (p<0.05) [see table 7.3 for test statistics]. This suggests that the Movement type had a significant effect on orientation estimation where the faster movement, running, had higher errors. The full set of ranks and test statistics for this Friedman test can be found in tables 7.7 and 7.8.

The additional Friedman test, to evaluate whether there was a significant change in orientation accuracy over time, showed that there was no statistically significant relationship between measurements across time points (Run0, Run10, Run15, Run20, Run25, Run30) [$\chi^2(17)$ =8.04, p =

0.24]. The full set of ranks and test statistics for this Friedman test can be found in tables 7.5 and 7.6.

Table 7.3: Wilcoxon Signed Rank Test results for the thigh in analysing the statistically significant relationship within Time and Movement Type.

Test Pair	Z_{Thigh}	p_{Thigh}
WalkStart-Run0	-2.81	0.005*
WalkStart-Run30	-2.30	0.02*
WalkStart-WalkEnd	-1.07	0.29
Run0-Run30	-1.63	0.10
Run0-WalkEnd	-3.10	0.002*
Run30-WalkEnd	-2.81	0.005*

Where p: significance p-value, Z: Wilcoxon Signed Rank test statistic, * denotes statistical significance.

Statistical Analysis Results of the Sacrum Segment data

Results from the Friedman test showed that there was a statistically significant relationship ($\chi^2(15)$ = 30.2, p = 1.0E-6) among the variables [Walkstart, Run0, Run30, WalkEnd]. Post-hoc analyses were carried out using Wilcoxon Signed Rank tests to examine every combination pair of the variables (Walkstart, Run0, Run30, WalkEnd). Each of the two pairs containing the same movement type were not significantly different [WalkStart - WalkEnd: (Z = -2.16, p = 0.03), Run0 - Run30: (Z = -0.9, p = 0.36)]. The 4 remaining pairs of differing movement type were all statistically significant (p<0.05), except for the pair WalkStart - Run30, as WalkStart was had higher errors which were more similar to Run30 (p=0.16) than WalkEnd (p=0.03) [see table 7.4 for test statistics]. This generally suggests that the Movement type had a significant effect on orientation estimation where the faster movement, running, had higher errors. The full set of ranks and test statistics for this Friedman test can be found in tables 7.7 and 7.8.

The additional Friedman test, to evaluate whether there was a significant change in orientation accuracy over time, showed that there was no statistically significant relationship between measurements across time points (Run0, Run10, Run15, Run20, Run25, Run30) [$\chi^2(15)$ =9.92 p = 0.13]. The full set of ranks and test statistics for this Friedman test can be found in tables 7.5 and 7.6.

Table 7.4: Wilcoxon Signed Rank Test results for the sacrum in analysing the statistically significant relationship within Time and Movement Type.

Test Pair	Z_{Sacrum}	p_{Sacrum}
WalkStart-Run0	-3.24	0.001*
WalkStart-Run30	-2.41	0.16
WalkStart-WalkEnd	-2.16	0.03*
Run0-Run30	-0.90	0.36
Run0-WalkEnd	-3.40	0.001*
Run30-WalkEnd	-3.40	0.001*

Where p: significance p-value, Z: Wilcoxon Signed Rank test statistic, * denotes statistical significance.

Table 7.5: Friedman test ranks for the tibia, thigh and sacrum segments in analysing orientation accuracy across Time.

Time	Mean Rank$_{Tibia}$	Mean Rank$_{Thigh}$	Mean Rank$_{Sacrum}$
Run0	3.95	3.65	4.27
Run5	4.45	4.00	4.33
Run10	4.25	3.75	4.20
Run15	3.80	5.00	4.10
Run20	3.65	4.29	4.87
Run25	3.95	4.24	2.60
Run30	3.95	3.06	3.63

.

Table 7.6: Friedman test statistics for the tibia, thigh and sacrum segments in analysing orientation accuracy across Time. It can be seen that there were no statistically significant changes in orientation accuracy over time across all segments ($p > 0.05$).

Segment	Chi-Square	p	N
Tibia	1.86	0.93	20
Thigh	8.04	0.24	17
Sacrum	9.92	0.13	15

Where p: significance p-value, N: Number of participant data in the analysis.

7.3.3 Discussion

In this research we captured inertial sensor data (accelerometer and gyroscope) on 20 participants walking and running on a treadmill for approximately 30-minutes. The inertial data was input to the Madgwick 6DOF filter to calculate tibia, thigh and sacrum segment orientation (on each segment axis). The Madgwick 6DOF filter does not take into account signal drift on the Z-axis (yaw angle) but corrects the drift on the X and Y axes (roll and pitch angles) using fusion of the accelerometer and

Table 7.7: Friedman test ranks for the tibia, thigh and sacrum segments for analysing the effect of Time and Movement Type.

Time	Mean Rank$_{Tibia}$	Mean Rank$_{Thigh}$	Mean Rank$_{Sacrum}$
WalkStart	2.10	2.18	2.13
Run0	3.00	3.24	1.20
Run30	3.10	2.88	3.60
WalkEnd	1.80	1.71	3.07

.

Table 7.8: Friedman test statistics for the tibia, thigh and sacrum segments in analysing the effect of Time and Movement Type. It can be seen that each segment had a statistically significant relationship ($p < 0.05$).

Segment	Chi-Square	p	N
Tibia	15.1	0.002*	20
Thigh	14.5	0.002*	17
Sacrum	30.2	1.0E-6*	15

. Where p: significance p-value, N: Number of participant data in the analysis, * denotes statistical significance.

gyroscope signals. A linear de-trending approach was used to locally remove drift from the selected section.

The primary aim of this study was to implement and determine the accuracy of the Madgwick's 6DOF orientation estimation filter [14], in order to assess the level of inaccuracies present during dynamic movement (running) over an extended time duration. The key factors of interest involved were the differing speeds of movement captured (walking and running) and the extended capture time (over a 30-minute period). The majority of studies which have evaluated human segment orientation/joint angles during fast dynamic movements, mainly relying on the use of a magnetometer. Studies which do not use a magnetometer tend to deal with the signal drift by using biomechanical constraints, joint centre optimisation or only calculate inclination angles (roll and pitch).

An additional aim was to develop a means to remove yaw orientation signal drift without the use of magnetometer data. A main finding showed that linear de-trending was capable of removing signal drift from yaw angles over segmented sections of comparison data. It was shown that linear de-trending performed better than a number of other drift removal techniques (using high pass filtering, mean trend removal, and polynomial trend removal) when these were applied on a subset of the data for comparison. Although time had no significant effect on the linear de-trending approach, on average

the magnitude of overall error in terms of absolute orientation on the Z-axis were still far too high to be considered clinically relevant or reliable. There does not appear to be studies which have investigated this approach as a means for 6DOF Z-axis orientation. This technique may prove useful in situations where absolute orientation is not required and relative orientation to the participant themselves may be acceptable. In a real-life data recording scenario without the Vicon reference, this Z-axis offset could be captured in an initialisation feature which would estimate the orientation offset of the Z-axis from zero, during a static stance pose, at the start of the data collection using the Madgwick filter.

The magnitude of average error across segment axes can be seen in table 7.1. Z-axis yaw estimation was better than gyroscope integration only, with the average Z-axis orientation across all segments approximately equal to 14.75 degrees (compared to 22.7 degrees on average across the X-axis). It can be seen in figures 7.5 (a) - (c) the linear de-trending had varied success at drift removal and alignment, leading to the overall average poor Z-axis orientation accuracy. The figures show instances where the linear de-trending worked extremely well, satisfactory, and very poorly. The success of the drift removal depended on the level of error in the local signal (as de-trending was performed on a segmented section basis). Visual investigation of the results led to instances seen in figure 7.5 (c) where the inertial sensor estimated orientation is so poor that the linear trend cannot be correctly identified. The errors seen in this figure are not signal drift. On examination of the data, it can be seen that these are instances where orientation has been estimated poorly due to the presence of fast movements, and an additional source of error. It was found that these erroneous estimates occurred directly after instances of data loss in the signal. This observation was seen throughout the dataset on varied participants and segments. It was not a single sensor specific issue. This orientation error appeared to be reliant on the magnitude of data loss, effecting the algorithm's ability to converge to the correct orientation.

With respect to the capture time, it was found that the linear de-trending removed the time-varying increase in orientation errors on the Z-axis, for both walking and running. No studies appear to have examined inertial sensor orientation accuracy for this extended length of time (approximately 30 minutes) during running, with 6DOF or 9DOF. Although it is well accepted that without compensation for signal drift, orientation errors grow unbounded with time [167]. Statistical analysis of the effect of time on running orientation RMSE showed there was no significant difference between the measured RMSE across time points. This suggests that the ability of the algorithm to compensate for drift was consistent throughout the experiment.

(a) Excellent drift removal and alignment.

(b) Satisfactory drift removal and alignment.

(c) Very poor drift removal and alignment.
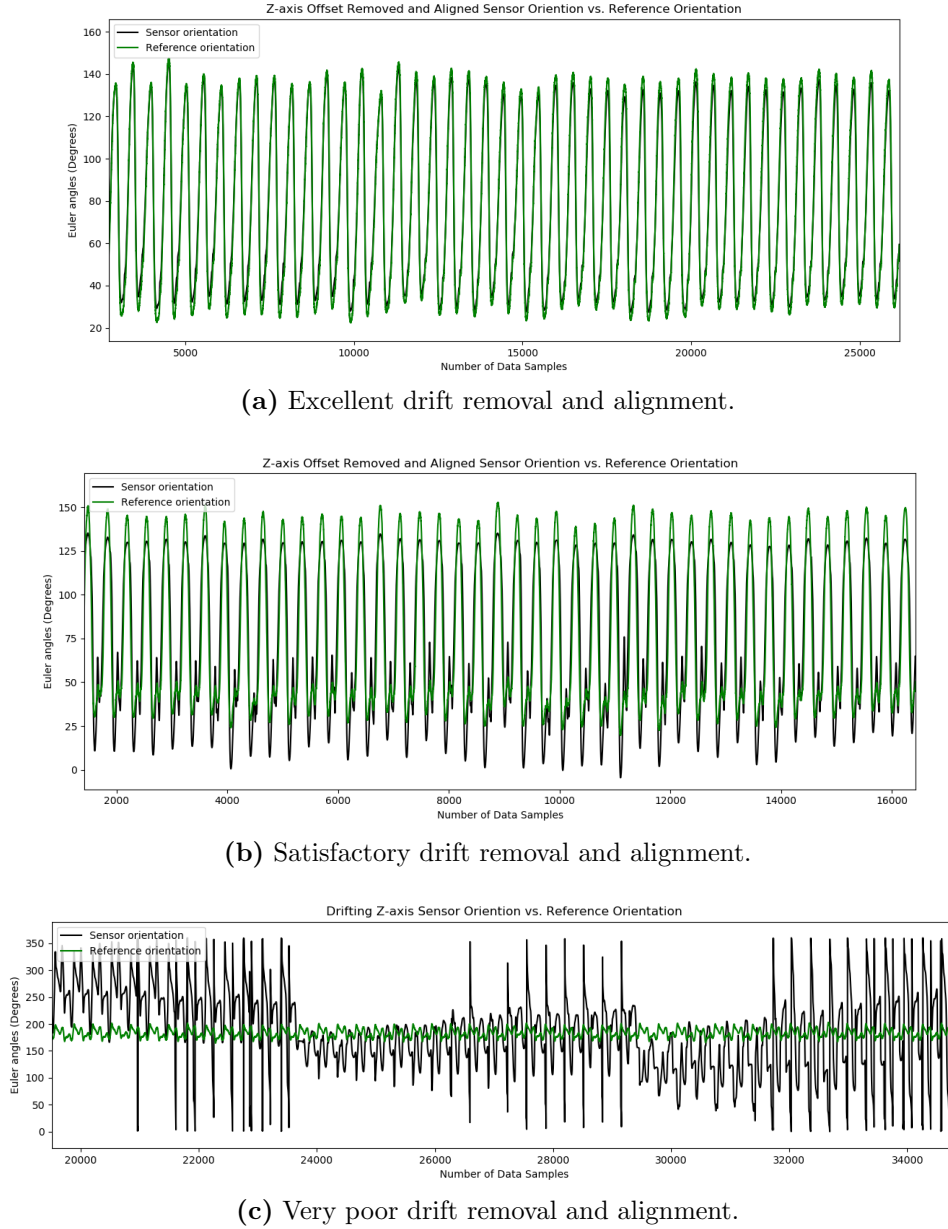
Figure 7.5: The linear de-trending had varied success at drift removal and alignment, leading to the overall average poor Z-axis orientation accuracy.

Regarding the effect of movement type, the results of walking and running RMSE across all segments show that the higher impacts and faster speeds associated with running produced higher orientation error. This has been directly examined in research where orientation during varied movement speeds have been compared [8, 164] and results reflect higher error associated with faster speeds regardless of orientation algorithm used [8], which agrees with the experimental results in this thesis. This is again reflected in the higher level of error on the X-axis, which was aligned with gravity and experienced the highest impact forces leading to noise and instability [366].

Fundamental to the Madgwick filter design is the need of the measured accelerometer vector to be compared with the gravity reference vector. In slow or quasi static movements, the measured accelerometer vector primarily measures gravity. In faster movements the accelerometer measures a combination of gravity and the movement acceleration (of the segment to which it is attached). This added error is an expected outcome, although the magnitude of this error was unclear due to the lack of research quantifying the level of inaccuracy during long durations of running. For this reason, we did not incorporate a mechanism to separate body accelerations and gravity in the measured acceleration data. Shull et al. [8] showed when measuring trunk orientation during walking and running that a 6DOF orientation estimation could outperform a 9DOF algorithm design in the presence of magnetic distortion (treadmill). Despite the lack of slow movements in this thesis, the 6DOF orientation results outlined in this chapter were still an improvement on calculations relying on gyroscope data only and did not suffer from local magnetic distortion.

Errors across all segment axes could also possibly be due to the high levels of data loss which occurred in this experiment, and the lack of beta parameter optimisation. The beta parameter value used in Study 1.1 was taken from Madgwick's original work [14], this is possibly a source of error as this parameter dictates the choices made by the algorithm to account for errors in the gyroscope orientation calculation. Tuning this parameter has been shown to have favourable improvements in orientation accuracy [19] and this will be explored in Study 1.3A of this thesis (Section 7.5) As the orientation errors reported in this thesis were rather large when compared to similar studies [8, 211], this approach is not sufficiently accurate to be used in for a human movement application.

7.3.4 Future Work

Further examination of data loss as a source of error should be explored due to the effect it may have had on these results, compounded by high impact accelerations and fast movements [366]. A solution to the issue of data loss should be a simple approach which does not add a computational burden to the orientation estimation process in order to be useful for real-time running applications. In addition, the possibility of improving the orientation accuracy through the optimisation of the beta parameter should be explored as this has proven to be successful in a number of studies examining other movements using both 6DOF and 9DOF estimation [12, 19, 366]. Both of these ideas will be explored in the following studies of this chapter.

7.3.5   Conclusion

In conclusion, the novel approach to estimate yaw orientation using linear de-trending and reference Z-offset values, proved to be an effective way to manage the existence of signal drift across time. However, Z-axis orientation errors were still too high to allow use of this approach for clinical assessments. All segment axes suffered from unacceptably large orientation errors. In order for orientation estimation to be used reliably for human movement analysis error <5 degrees would be more acceptable [8, 211, 367, 368]. The errors present in the data were due in-part to running being a high impact/fast movement, but additionally the magnitude of these errors may not have been related to high impacts/high speed alone. Further investigation pointed to the presence of data loss and a lack of algorithm optimisation to have added to the level of error.

7.4   Study 1.2: Addressing Sensor Data loss

In the previous study (Section 7.3), our results showed that erroneous orientation data occurred around instances of data loss. As the orientation calculation [14] relies on accurate sequential values to converge to an appropriate estimate, there was a temporary detrimental effect on the orientation estimate around points of data loss (until the algorithm had processed enough data after this point to be able to re-converge to the correct orientation). Though this effect only affected a window of data after the data loss, it occurred enough times to reasonably distort the dataset accuracy, creating long periods of inaccurate data. Xiao et al. [373] found that packet-loss rates in wireless inertial sensor systems can amplify pose-tracking errors by as much as 39%. This is particularly important if this occurs near an event of interest in the data (i.e., segment/joint orientation at initial contact). The initial work which developed the now commonly used Madgwick Filter [14] does not mention missing data, making it difficult to understand how the filter manages missing data, if at all.

Data loss is a phenomenon which affects all sensors [374], which includes inertial sensors [375, 376] and is not unique to the inertial sensors (Shimmer3 IMU) used in this thesis. To show this, experimental testing was carried out by (as part of this PhD) on three separate brands of inertial sensors, the Shimmer3 IMUs (Shimmer Sensing, Dublin), IMeasureU IMU sensors (Vicon, UK) and NGIMU sensors (X-IO Technologies, UK). The tests across all the sensors showed that each IMU suffered from data loss, with the Shimmer sensor having the least amount (this experiment was not included in the thesis).

The data loss may occur for several reasons: data transmission, battery failure, sensor malfunction [374, 377]. A sensor's internal processor can sometimes miss samples to be written to the SD card when the sampling rate and number of sensors recorded are high, increasing the amount of data to be processed. This is relevant because not only sensor applications that rely on a network or Bluetooth connection can suffer from data loss. The result of this technical failure is that all the inertial sensor data (i.e. accelerometer, gyroscope) is lost for a particular time point during data collection.

Managing missing time-series data is a well-researched area with many possible approaches such as ignoring or deletion of missing samples (deletion is not appropriate for real-time applications), data imputation methods which can fill data in a univariate sense using interpolation of linear, splines or polynomial curves, mean value imputation, last observation carried forward, K-NN or multiple imputation. There are also model based approaches which rely on the observed data to make estimates about missing values. This could be done using Maximum Likelihood Estimation or pattern matching models [377, 378]. In addition to these, recent approaches include machine learning and deep learning solutions, which either use the observed data to be able to predict missing data [379, 380] or create their models to be able to deal with missing data [374, 381].

With respect to missing data occurring in inertial sensor data and specifically missing inertial data, which is used to calculate orientation, there is little research showing what methods are typically applied. To date it appears only two studies Kim et al. [382] and Xiao et al. [373] have focused solely on examining the reconstruction of missing inertial data, and one study Lin et al. [380] on the reconstruction of orientation data (Euler angles); more research is clearly needed.

Aiming to tackle missing inertial data, Kim et al. [382] compared a number of approaches (cubic spline, empirical mode decomposition, and auto-regressive models) to fill specifically gyroscope inertial data and found a combination of empirical mode decomposition and an auto-regressive model to perform the best data reconstruction. Xiao et al. [373] used an RNN in order to deal with missing inertial data and reduce pose tracking (orientation) errors. These approaches can be thought of as dealing with missing data **before calculating orientation** which we will further refer to as 'pre-filling'. Pre-filling is defined as applying a pre-processing gap filling routine to the raw inertial data, tackling missing data in the inertial signals themselves before orientation is calculated [383]

Lin et al. [380] trained a neural network on observed Euler angle data in order to build a model to predict the Euler angles when gaps occurred. Lin et al. [380] was the only study found which handled missing data **after orientation was calculated** (via Euler angles) which we will further refer to as

'post-filling'. Post-filling is defined as applying a post-processing gap filling routine to the calculated orientation data (Euler angles of segment orientation or joint angle), which is tackling missing data after orientation is calculated. No studies have been found which compare the difference between **filling the missing data before calculating orientation** (pre-filling) and **filling the missing data after calculating orientation** (post-filling).

Other studies which directly examined filling missing sensor data in healthcare data based applications examined heart rate sensors using RNN [375], neural networks [376], K-nearest neighbour [376], multiple imputations [377], and a range of simple approaches: linear [376, 384], last observation carried forward [376, 384], quadratic [384], cubic [384]. Notice all studies except have been published recently, indicating a growth in the interest to tackle this problem.

As the research directly examining the imputation of missing data in inertial signals was limited, the following refers to studies which may have dealt with missing data indirectly and did not examine the impact this had on data. The primary approach appears to be to simply ignore the fact that there may be missing inertial data due to the lack of research to tackle this issue. A second approach may be to reduce the sampling rate in order to alleviate the effect of the missing data [371], although this approach may still pose an issue when gap sizes of missing data are quite large [385]. Despite the fact that these two approaches may be suitable for some applications, problems occur where the target application may require high sampling rates and rely on the analysis of particular events occurring at defined time points in the data. The analysis of events in this way is commonly undertaken in the area of biomechanics, for example when identifying joint angles at the time of initial foot contact with the ground during walking or running [386, 387].

A third approach is to fill the gaps with a gap filling method [315]. A number of studies have used various approaches to fill inertial data for the purpose of data re-sampling [370, 371, 388, 389]. Kanzler et al. [390], Dorschky et al. [370], Munoz et al. [371] and Scheurer et al. [315] used linear interpolation on their inertial data before inputting the data to deep learning frameworks. Hoflinger et al. [388], Neville et al. [389] and Kanzler et al. [390] used cubic spline interpolation on inertial data to estimate sensor temperature drift, postural stability and to compensate for velocity drift respectively. Of these studies, only Munoz [371] and Scheurer et al. [315] mentioned that the inertial data linear interpolation was to compensate for data loss.

To summarise, studies which use inertial sensors primarily ignore the missing data, re-sample the data in post-processing, or use a simple gap filling approach (linear or cubic spline). There is

not much research focusing solely on recovering missing inertial data. Studies which examined this found complex approaches of deep neural networks [373, 380] and auto-regressive/empirical mode decomposition models [382] to be useful. With regard to Kim et al. [382], the approach used must be performed offline in post-processing as it uses data before and after the gap of missing data in order to do imputation. This is not suitable for real-time applications and additionally this can be complex and inefficient to calculate [380]. Overly complex and complicated approaches tend to hinder applicability for real world scenarios [391], and when the occurrence of data loss is fairly common a suitable compromise between computational speed and quality of results has to be made [392]. Therefore, it is worth trying a simpler univariate approach such as linear or spline interpolation as used for inertial data re-sampling. Though they typically perform better on smaller gaps, these methods are simple and efficient with low calculating complexity [381, 393], and have been shown on occasion to outperform more complicated gap-filling methods [376, 394]. These may be more suited to real-time applications, which was the aim of this work, so that it could be used in the second main study of the thesis (Section 8).

Due to the fact that sensors lose varying amounts of consecutive samples, it may be important to deal with these varying gap sizes in different ways i.e., smaller gap sizes may be easier to fill accurately than the larger gap sizes [385, 395]. It has already been shown that simple interpolation methods can work quite well on small gap sizes [394, 396] so it will be investigated whether filling some of the missing data (up to a set gap size) will improve our orientation estimation. This could be very effective at helping the Madgwick filter converge back to the true orientation quicker after periods of data loss in real-time, and to reduce the impact on subsequent estimates. The experiments in this section (Study 1.2) investigate the effect of varying gap sizes through the use of gaps size thresholds, up to which missing samples would be filled in the inertial data. This was to determine whether gaps need to be treated differently based on the number of samples lost and whether a cut-off point existed for filling gaps using the methods explored. Gap size thresholds of: 0 missing samples, <10 missing samples, <20 missing samples, <40 missing samples, <80 missing samples, <100 missing samples, <400 missing samples, were decided on experimentally (Section 11.1).

Broadly, this Section 7.4 aims to evaluate the performance of creating a gap filling pre-processing pipeline of pre-filling and post-filling for improved orientation accuracy in limiting the negative effects of data loss. This was done by investigating whether filling some of the missing inertial data (pre-filling up to a fixed gap size threshold) resulted in more accurate orientation data than not filling any gaps

in inertial data. This can be summarised as comparing the effectiveness of pre-filling (filling gaps in inertial data before orientation calculation) and post-filling (filling gaps in orientation data) the missing data. A sub-aim during the experiment was to determine whether simple interpolation methods, **linear interpolation** (as used in Study 1.1) and 'Last Observation Carried Forward' can sufficiently reconstruct missing inertial data for the specific purpose of orientation estimation, and which of the two gap filling approaches performed better. 'Last Observation Carried Forward' approach which will be referred to as **'previous sample filling'**.

Therefore, the effect of three factors on the orientation data were examined: Gap size threshold, pre-fill method (linear interpolation or previous sample filling) and post-fill method (linear interpolation or previous sample filling). The experiments investigated whether it was better to solely 'pre-fill' the missing data, i.e., fill all missing samples in inertial data (Threshold size <400), solely 'post-fill' the missing data, i.e., fill all missing samples in the orientation data (Threshold size 0), or it was better to both 'pre-fill' and 'post-fill', by pre-filling missing samples up to the gap size threshold (<10, <20, <40, <80, <100) and post-filling remaining gaps after orientation was calculated.

7.4.1 Methodology

7.4.1.1 Data Processing

The first experiment involved taking the inertial data of 5 participants from the dataset (used in the previous section) and filling all missing data which occurred in the inertial data. This was repeated for several gap filling approaches, and orientation was calculated with the filled inertial data. The root mean square error (RMSE) between the sensor data and the Vicon markers was calculated (as described in the previous section) and this was compared to the RMSE of orientation calculated without filling the missing inertial data samples for each participant. The gap filling methods used were: no fill, previous sample filling (same as last observation carried forward), linear interpolation, quadratic interpolation, cubic spline interpolation, quartic spline interpolation and quintic spline interpolation.

The methods of gap filling were chosen as they are simple and efficient [393], due to their appropriate nature for a real-time application with low computational overhead [381], frequent use in gap filling of other time-series data [376, 377, 384] and as they have been shown on occasion to outperform more complicated gap-filling methods [376, 394]. The main issues mentioned with these simple approaches are that they don't take into account any assumptions about the variability of the

data to be filled [381, 397], and so may not be appropriate for all applications as they tend to underestimate the missing values [398]. Although using a method like linear interpolation (which is bounded by its start and end points) over the spline approaches means at least there their will not be any overshoot within the estimates, which can produce uncharacteristically high values for the data [398]. Although more complication solutions might necessarily be better suited to these problems due to this variability of the data from person to person, model-based solution may struggle with being able to generalise [376, 377] without some complementary correlated information, or individualised subject based models. Issues such as the overhead associated with multi-sensor/sensor axis prediction synchronously [373] and using previously predicted values to impute preceding gaps [380] complicate these matters further. Overly complex and complicated approaches tend to hinder applicability for real world scenarios [391], and when the occurrence of data loss is fairly common a suitable compromise between computational speed and quality of results has to be made [392]. Therefore, it is worth trying simpler single imputation approaches. Additionally, simple methods have been shown to be provide reasonable estimates for small gap sizes [393, 394, 396, 397, 399, 400]. This is relevant as the results from a preliminary experiment (section 11.1) showed that the magnitude of data loss had a frequent occurrence of small gaps (i.e., many instances of < 10 consecutive missing samples). Data loss did also occur at higher magnitudes.

The results of the preliminary gap filling analysis can be seen in figure 7.6, where the best gap filling methods were: not filling missing data, filling all missing inertial data with a linear interpolation, or filling all missing inertial data with previous sample filling. In fact, the gap filling methods performed equally to not filling the inertial data; it is expected that this is due to the inclusion of larger gaps being filled with the interpolation which drags down the average accuracy. This result, paired with the fact that most instances of data loss were occurring in 0-10 consecutive missing samples, lead us to question whether filling these small gaps sufficiently could boost the accuracy of the orientation estimation overall. This section investigated using gap size thresholds, up to which missing samples were filled in the inertial data (pre-fill), then orientation was calculated, and any remaining gaps were filled in the orientation data (post-fill).

Using the results from the preliminary experiment (see Section 11.1), the effect of data loss was evaluated on the data of 10 randomly selected subjects from the common dataset for this Chapter (dataset described in Section 6.2). Using the knowledge that the average maximum data loss for our data was in the range of 80-100 consecutive lost samples (Section 11.1), gap size thresholds were
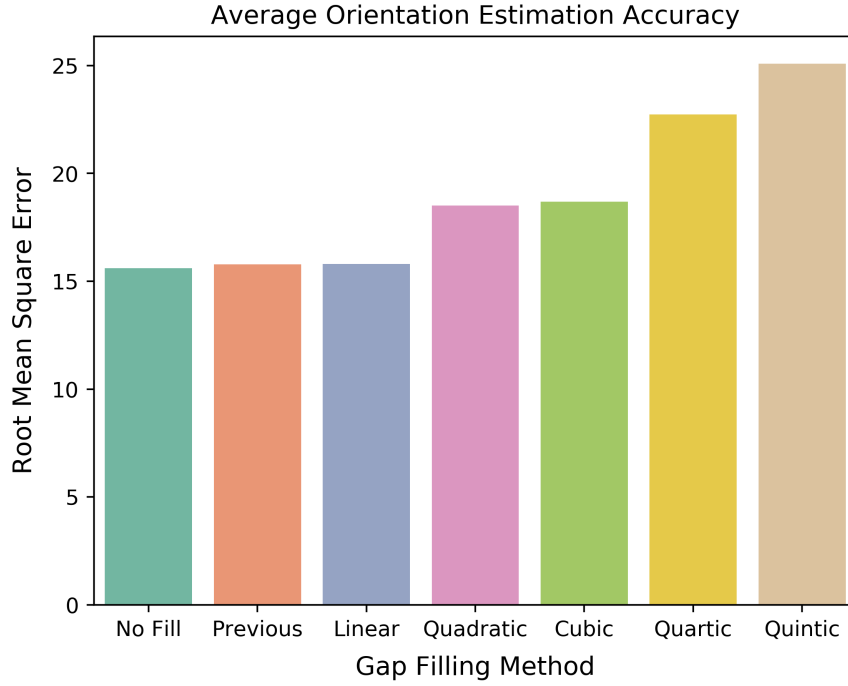
Figure 7.6: Inertial data gap filling analysis of a number of approaches. It can be seen that linear interpolation and previous sample filling perform the best.

chosen, up to which gaps in the inertial data would be filled. The thresholds were set as: 0, <10, <20, <40, <80, <100, <400, where <20 means that any group of consecutive missing samples smaller than 20 samples in the inertial data would be gap filled, and any larger groups of consecutive missing samples would be ignored in the pre-filling of the data. This experiment involved pre-processing the inertial data with these thresholds by filling consecutive missing samples below the threshold, where 0 means that no missing samples in the inertial data were pre-filled, and <400 means that all missing samples in the inertial data were pre-filled. This was carried out on the data of each segment of the 10 randomly selected subjects, and orientation was calculated with the pre-filled data for each set threshold. Due to the issue of signal drift which would accumulate in the Z-axis (due to using a 6-DOF orientation estimation approach), and as these experiments are in part investigating the effectiveness of the pre-processing pre-fill step, including the Z-axis would add an additional source of error to the results. The behaviour of data loss is inherently random and not behaviour specific to each axis and so the Z-axis was not included in the analysis. By excluding the Z-axis any errors present after the orientation step are due to the threshold sizes and gap filling method alone.

Orientation was calculated with the pre-filled inertial data using the same Madgwick filter [14] as described in the Methodology of Study 1.1 (Section 7.3). Orientation calculation was carried out

twice, once with the gap-filling method set to linear interpolation in the pre-fill step, and again with the gap-filling method set to previous sample filling to investigate which was more appropriate to provide more accurate pre-filled data. For each segment, a set of 14 orientation data were calculated (Linear pre-fill with each Threshold size (0, <10, <20, <40, <80, <100, <400) and previous sample pre-fill with each Threshold size (0, <10, <20, <40, <80, <100, <400).

Once the orientation data were calculated, the remaining missing samples were post-filled in order to keep the time-aligned synchronisation with the reference Vicon data from the dataset. This was carried out twice for each set of pre-filled data (i.e., data with linear pre-fill and data with previous sample pre-fill), once with the post-filling method set to linear interpolation, and again with the post-filling method set to previous sample filling to investigate which was more appropriate to provide accurate post-filled data. This methodology flow is described in figure 7.7 for clarity. Each subject had 9 data comparison sections recorded during the experiment (walk, 0-minute run, 5-minute run, ..., 30-minute run, walk), where the orientation estimates from inertial data could be compared with the Vicon reference orientation data using RMSE. RMSE was averaged across all comparison sections (walk0, run0,..., run30,walk30) of each set of orientation data under each specific condition, totalling 28 averages for each axis orientation (X and Y) representing the 28 condition combinations (7 threshold sizes: 0, <10, <20, <40, <80, <100, <400, 2 pre-fill types: linear and previous, and 2 post-fill types: linear and previous) for each participant.

7.4.1.2 Data Analysis

RMSE was used to compare the outputs of each combination of the 7 threshold sizes (0, <10, <20, <40, <80, <100, <400), 2 pre-fill types (linear and previous) and 2 post-fill types (linear and previous) to the reference Vicon data. The distribution of the data was assessed by Shapiro-Wilk's test for normality, and it was found that the data was non-normally distributed ($p < 0.05$). It has been shown in cases on non-normal data with large sample sizes that the family of t and F parametric tests are robust [401, 402]. Statistical analysis was carried out using a 3-way repeated measure analysis of variance (ANOVA), with RMSE as the dependent variable and the threshold size, pre-fill type, and post-fill type were the three independent factors. The ANOVA was carried out to get an insight into how these factors impacted the accuracy of the orientation data. Data were analysed at a segment and axis level, with a separate ANOVA for each: Tibia X-axis, Tibia Y-axis, Thigh X-axis, Thigh Y-axis, Sacrum X-axis and Sacrum Y-axis. Where a significant interaction effect was found involving
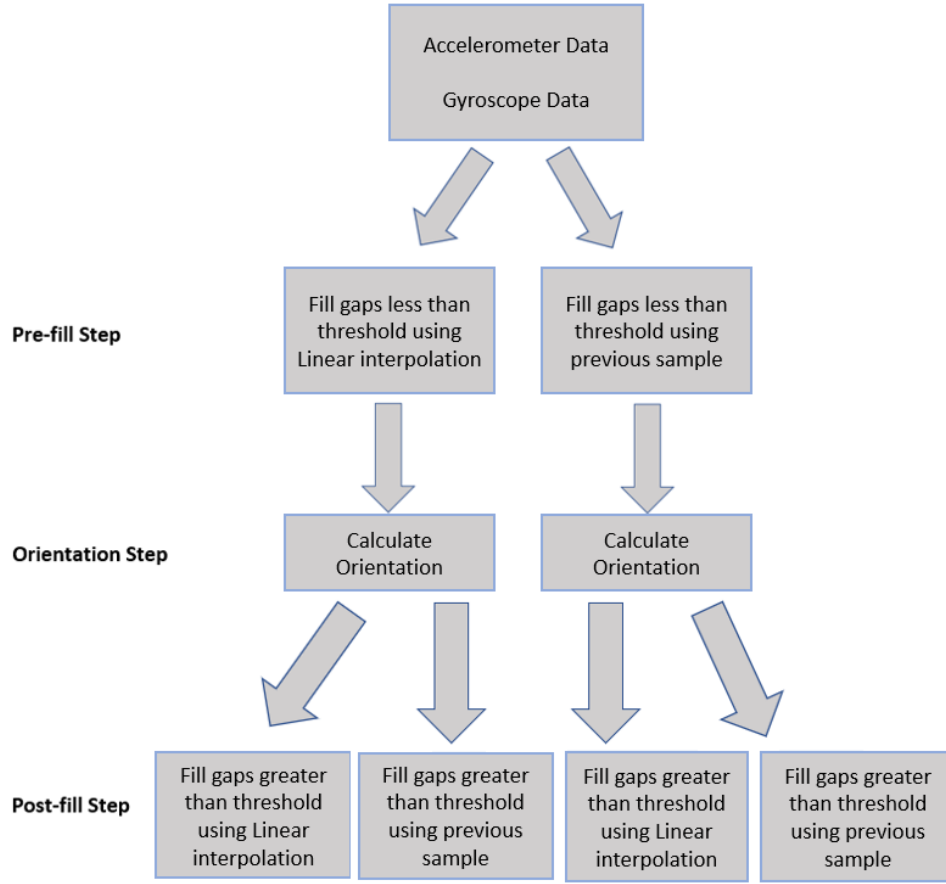
Figure 7.7: Methodology flow carried out for each threshold size (0, 10, 20, 40, 80, 100, 400).

solely threshold size and pre-fill type or solely threshold size and post-fill type, paired t-tests were carried out as post-hoc analyses. The level of significance was set at p < 0.05. The effect size was measured using (partial) eta squared (η^2), where $\eta^2 = 0.01$ indicated a small effect, $\eta^2 = 0.06$ indicated a medium effect and $\eta^2 = 0.14$ indicated a large effect size. Where Mauchly's test of Sphericity was violated, Greenhouse-Geisser correction of the degrees of freedom was used. The statistical analysis was completed using IBM SPSS Statistics 24. Data analysis was repeated with the data transformed to a normal distribution using $\log_{10}$ (not included in thesis) to ensure the reliability of the results. The repeated results were in line with the parametric results included in this section.

7.4.2   Results

To evaluate whether a gap filling pre-processing pipeline of pre-filling and post-filling missing samples of fixed gap-size could be effective on improving orientation accuracy, statistical analysis was carried out to examine the effect of threshold size, pre-fill method and post-fill on the accuracy of the segment orientation data. The results of the three-way repeated measures ANOVA for each segment axis

are shown in tables 7.9, 7.10, 7.15, 7.16, 7.17 and 7.18. T-test p values were adjusted for multiple comparisons (n=7) using Bonferroni correction. The results of the paired t-tests where appropriate are shown in tables 7.11, 7.12, 7.13, 7.14. Some of the results of the t-tests (e.g., table 7.11) show no results for Thresholds <10, <20, <80, <100, <400, as correlation values could not be computed in SPSS when the standard error of the difference is zero.

Statistical Analysis Results of the Tibia Segment data

The ANOVA tests for both the Tibia X and Y axes had similar results. The ANOVA tests showed a significant interaction between threshold size and post-fill type (X-axis: $F_{(1,89)}=8.09$, $p<0.05$, Y-axis: $F_{(1,89)}=17.41$, $p <0.05$) and also that threshold size and post-fill type both had a significant main effect ($p< 0.05$) (see tables 7.9 and 7.10). Post-hoc analyses of the interaction effect were carried out across pairs of differing Post fill type and Threshold size. There was a statistically significant increase in error associated with the linear interpolation of Threshold size 0, for both axes (x-axis adjusted $p =0.035$, y-axis adjusted p=3.7E-4). Although the increase in error was statistically significant, the mean RMSE difference was not a practically significant result where Linear interpolation produced marginally greater error than previous sample fill with a mean RMSE difference of approximately 0.03 degrees for both axes (see tables 7.11 and 7.12). The plots of the means (figure 7.8) show that the difference in magnitudes of post-fill type across gap sizes is inconsequential, which is reflected in the difference between linear interpolation (X-axis mean=20.76, SD=13.81 and Y-axis mean=8.51, SD=5.760) and previous sample post-filling (X-axis mean=20.73, SD=13.79 and Y-axis mean=8.48, SD=5.759) for threshold size 0.

Table 7.9: Three-way repeated measures ANOVA results showing Threshold size and Post-fill type as a significant main effect, and a significant interaction between Threshold-size and Post-fill type for Tibia X-axis data.

Source	df	MS	F	p	Effect Size
Threshold Size	1.53	2220	3.82	0.04*	0.04
Pre-fill type	1.00	0.27	1.51	0.22	0.02
Post-fill type	1.00	0.01	8.56	4.0E-3*	0.09
Threshold X Pre-fill	1.40	0.06	0.39	0.61	4.0E-3
Threshold X Post-fill	1.00	0.07	8.09	6.0E-3*	0.08
Pre-fill X Post-fill	1.00	1.3E-10	0.56	0.46	6.0E-3
ThresholdXPre-fillXPost-fill	2.04	1.0E-10	0.37	0.67	4.0E-3

df = Degrees of Freedom, MS = Mean Squared value, F = f value, p = significance level. Asterisks show significance.

Table 7.10: Three-way repeated measures ANOVA results showing Threshold size and Post-fill type as a significant main effect, and a significant interaction between Threshold-size and Post-fill type for Tibia Y data

Source	df	MS	F	p	Effect Size
Threshold Size	1.94	363	4.70	0.01*	0.05
Pre-fill type	1.00	0.002	0.04	0.85	4.1E-4
Post-fill type	1.00	0.005	7.30	0.01*	0.08
Threshold X Pre-fill	1.36	0.01	0.17	0.76	0.002
Threshold X Post-fill	1.23	0.03	17.41	1.6E-5*	0.16
Pre-fill X Post-fill	1.00	1.8E-4	0.92	0.34	0.01
ThresholdXPre-fillXPost-fill	1.01	2.9E-10	1.08	0.30	0.12

df = Degrees of Freedom, MS = Mean Squared value, F = f value, p = significance level. Asterisks show significance.

Table 7.11: T-test analysis between orientation RMSE pairs examining the threshold sizeXPost-fill type interaction for the Tibia X-axis data.

Threshold Size	Post-fill Type	Mean Difference	SD	t	p
0	Linear	-	-	-	-
0	Previous	0.03	0.10	2.9	0.005*
10	Linear	-	-	-	-
10	Previous	3.3E-4	0.003	1.14	0.26
20	Linear	-	-	-	-
20	Previous	3.3E-4	0.002	1.75	0.08
40	Linear	-	-	-	-
40	Previous	1.1E-4	0.001	1.00	0.32
80	Linear	-	-	-	-
80	Previous	1.1E-4	0.001	1.00	0.32
100	Linear	-	-	-	-
100	Previous	1.1E-4	0.001	1.00	0.32

SD = Standard Deviation difference, p = significance level. Asterisks show significance.

Statistical Analysis Results of the Sacrum Segment data

The ANOVA tests for both the Sacrum X and Y axes had similar results. The ANOVA tests showed a significant interaction between threshold size and post-fill type (X-axis: $F(1,89)= 33.20$, $p<0.05$, Y-axis: $F(1,89)= 27.18$, $p<0.05$) [see tables 7.15 and 7.16]. Post-hoc analyses of the interaction effect were carried out across pairs of differing post fill type and Threshold size.

Table 7.12: T-test analysis between orientation RMSE pairs examining the threshold sizeXPost-fill type interaction for the Tibia Y-axis data.

Threshold Size	Post-fill Type	Mean Difference	SD	t	p
0	Linear	-	-	-	-
0	Previous	0.022	0.05	4.25	5.3E-5*
10	Linear	-	-	-	-
10	Previous	0.001	1.5E-3	2.16	0.033*
20	Linear	-	-	-	-
20	Previous	5.4E-4	6.0E-5	1.00	0.32
40	Linear	-	-	-	-
40	Previous	1.1E-4	7.4E-4	1.42	0.16
80	Linear	-	-	-	-
80	Previous	-0.001	0.013	-1.00	0.32
100	Linear	-	-	-	-
100	Previous	-0.001	0.013	-1.00	0.32
400	Linear	-	-	-	-
400	Previous	-0.001	0.013	-1.00	0.32

SD = Standard Deviation difference, p = significance level. Asterisks show significance.

Table 7.13: T-test analysis between orientation RMSE pairs examining the threshold sizeXPost-fill type interaction for the Sacrum X data.

Threshold Size	Post-fill Type	Mean Difference	Std	t	p
0	Linear	-	-	-	-
0	Previous	-0.04	0.06	-5.8	9.4E-8*
10	Linear	-	-	-	-
10	Previous	-8.4E-4	7.1E-3	-1.1	0.27
20	Linear	-	-	-	-
20	Previous	-2.0E-5	1.3E-4	-1.4	0.16
40	Linear	-	-	-	-
40	Previous	-4.0E-6	6.0E-5	-0.59	0.55

Std Deviation = Standard Deviation difference, p = significance level. Asterisks show significance.

There was a statistically significant increase in error associated with the previous interpolation of Threshold size 0, for both axes (X-axis adjusted p =6.5E-7, and y-axis adjusted p=7.0E-6). Similarly to the tibia, although the increase in error was statistically significant, the mean RMSE difference was not a practically significant result where Previous interpolation produced marginally greater error than Linear interpolation with a mean RMSE difference < 0.04 degrees for both axes (see tables 7.13

and 7.14). The plots of the means (figure 7.9) show that the difference in magnitudes of post-fill types across gap sizes is unimportant, which can be seen in the difference between linear interpolation (X-axis mean=21.39, SD=28.59 and Y-axis mean=7.377, SD=3.77) and previous sample post-filling (X-axis mean=21.42, SD=28.61 and Y-axis mean=7.382, SD=3.77) for threshold size 0.

Table 7.14: T-test analysis between orientation RMSE pairs examining the threshold size X Post-fill type interaction for the Sacrum Y data.

Threshold Size	Post-fill Type	Mean Difference	Std	t	p
0	Linear	-	-	-	-
0	Previous	-4.4E-4	8.0E-4	-5.2	1.0E-6*

Std Deviation: Standard Deviation difference, p: significance level. Asterisks show significance.

Table 7.15: Three-way repeated measures ANOVA results showing Threshold size and Post-fill type as a significant main effect, and a significant interaction between Threshold-size and Post-fill type for Sacrum X-axis data

Source	df	MS	F	p	Effect Size
Threshold Size	1.8	17	0.36	0.68	0.004
Pre-fill type	1.00	0.55	0.54	0.46	0.006
Post-fill type	1.00	0.02	34.02	8.4E-8*	0.28
Threshold X Pre-fill	1.09	1.27	1.56	0.22	0.02
Threshold X Post-fill	1.03	0.10	33.2	8.2E-8*	0.27
Pre-fill X Post-fill	1.00	5.0E-6	1.00	0.32	0.011
Threshold X Pre-fill X Post-fill	1.00	3.0E-5	0.99	0.32	0.011

df = Degrees of Freedom, MS = Mean Squared value, F = f value, p = significance level. Asterisks show significance.

Statistical Analysis Results of the Thigh Segment data

The ANOVA tests for both the Thigh X and Y axes had similar results. The ANOVA tests showed no significant interaction or main effects for threshold size, pre-fill type and post-fill type (see tables 7.17 and 7.18). This can be explained by experiments which were included in the Appendix (see Section 11.1), where the data loss experienced was analysed across sensors. It can be see that the magnitudes of data loss experienced by the thigh was less severe compared to the tibia and sacrum (see average maximum gap sizes experiences across segments in figures 11.4, 11.5 and 11.6), and although large gap sizes did occur, on average the thigh data experienced at most losses of 30-34 consecutive samples (figure 11.9). This means that when the data were analysed in terms of fixed gap sizes (0, 10, 20, 40, 80, 100, 400), there was no major difference in investigating gap sizes >40 for the thigh with few gaps

reaching the >40 consecutive samples criteria, and so <80, <100, <400 produced close results to <40.

Table 7.16: Three-way repeated measures ANOVA results showing Threshold size and Post-fill type as a significant main effect, and a significant interaction between Threshold-size and Post-fill type for Sacrum Y-axis data

Source	df	MS	F	p	Effect Size
Threshold Size	1.6	4	5.00	0.01*	0.05
Pre-fill type	1.00	0.03	2.48	0.12	0.03
Post-fill type	1.00	2.5E-4	27.18	1.0E-6*	0.23
Threshold X Pre-fill	1.19	0.004	0.74	0.41	0.01
Threshold X Post-fill	1.00	0.002	27.18	1.0E-6*	0.23
Pre-fill X Post-fill	1.00	0	-	-	-
Threshold X Pre-fill X Post-fill	1.00	0	-	-	-

df = Degrees of Freedom, MS = Mean Squared value, F = f value, p = significance level. Asterisks show significance.

Table 7.17: Three-way repeated measures ANOVA results showing no significant main effect or significant interaction between Threshold-size, Post-fill type and Pre-fill type for Thigh X-axis data.

Source	df	MS	F	p	Effect Size
Threshold Size	1.39	0.03	3.07	0.07	0.03
Pre-fill type	1.00	6.6E-5	3.35	0.07	0.04
Post-fill type	1.00	3.7E-9	2.37	0.13	0.03
Threshold X Pre-fill	1.41	1.1E-5	2.32	0.12	0.03
Threshold X Post-fill	1.39	3.8E-9	0.70	0.45	0.01
Pre-fill X Post-fill	1.00	1.6E-10	0.30	0.59	0.003
Threshold X Pre-fill X Post-fill	1.04	4.2E-10	0.31	0.59	0.003

df = Degrees of Freedom, MS = Mean Squared value, F = f value, p = significance level. Asterisks show significance.

Table 7.18: Three-way repeated measures ANOVA results showing Threshold size and Pre-fill type as a significant main effect, and a significant interaction between Threshold-size and Pre-fill type for Thigh Y data

Source	df	MS	F	p	Effect Size
Threshold Size	1.29	4.09	2.48	0.11	0.03
Pre-fill type	1.00	0.01	3.50	0.07	0.04
Post-fill type	1.00	1.5E-7	1.00	0.32	0.01
Threshold X Pre-fill	1.57	0.002	2.49	0.10	0.03
Threshold X Post-fill	1.00	9.4E-7	1.00	0.32	0.01
Pre-fill X Post-fill	1.00	1.5E-7	1.00	0.32	0.01
Threshold X Pre-fill X Post-fill	1.00	9.4E-7	1.00	0.32	0.01

df = Degrees of Freedom, MS = Mean Squared value, F = f value, p = significance level. Asterisks show significance.

(a) X-axis tibia data
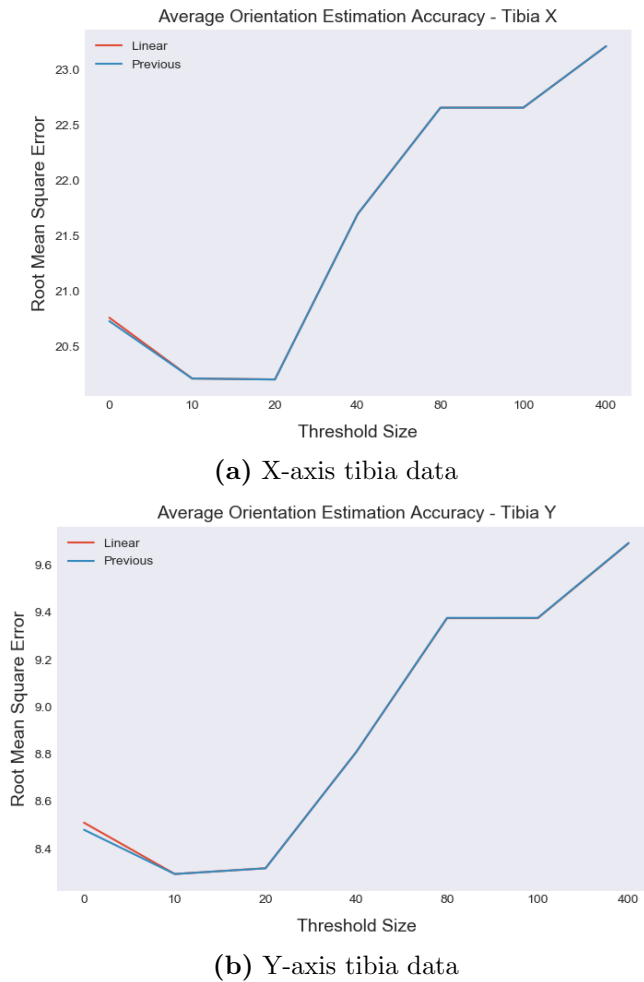
(b) Y-axis tibia data

Figure 7.8: Average orientation RMSE across threshold sizes for tibia data where gaps at each threshold were post-filled using linear interpolation (blue) and previous sample filling (red).
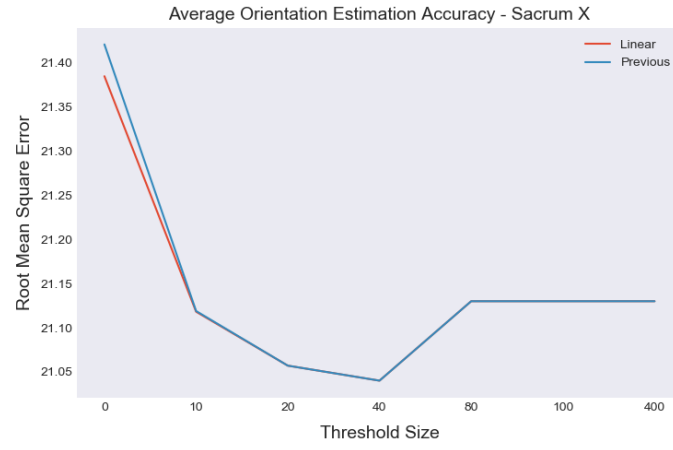
(a) X-axis sacrum data

(b) Y-axis sacrum data

Figure 7.9: Average orientation RMSE across threshold sizes for sacrum data where gaps at each threshold were post-filled using linear interpolation (blue) and previous sample filling (red).

(a) X-axis thigh data

(b) Y-axis thigh data
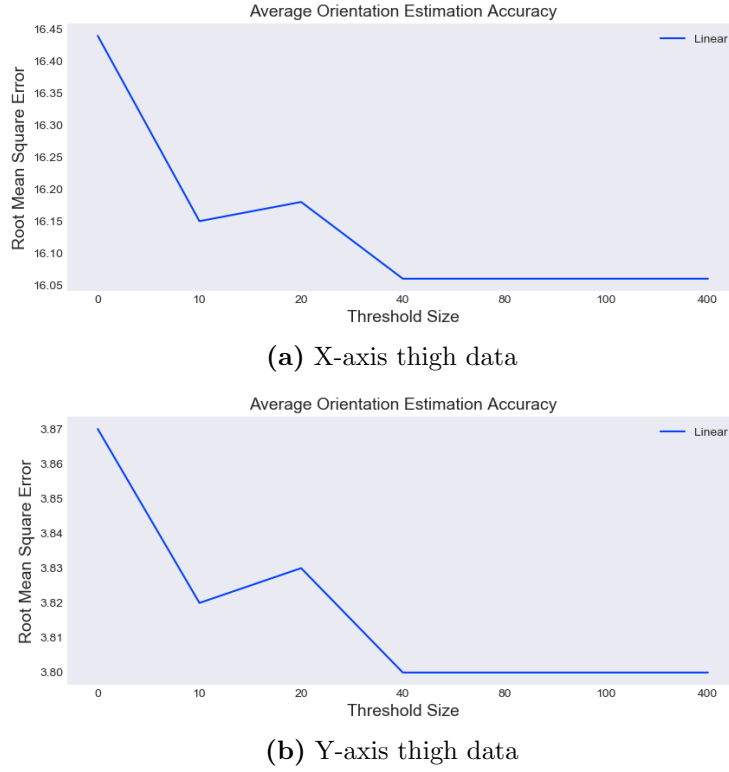
Figure 7.10: Average orientation RMSE across threshold sizes for thigh data where gaps at each threshold were post-filled using linear interpolation.

7.4.3 Discussion

The primary aim of this section (Study 1.2) was to investigate whether a gap-filling pre-processing pipeline of filling missing inertial data of varying sizes would lead to more accurate orientation data, and whether linear interpolation or previous sample filling was more effective at accurately filling the data for the purpose of calculating orientation. The methods of gap filling were chosen as they are simple and efficient [393], due to their appropriate nature for a real-time application with low computational overhead [381], frequent use in gap filling of other time-series data [371, 384, 403] and as they have been shown on occasion to outperform more complicated gap-filling methods [376, 394].

Preliminary testing showed that there were a large number of small gaps in the data, and experiments examined whether a set threshold gap size (up to which missing data would be filled) would improve performance. In theory this could help the Madgwick filter converge back quicker to the true orientation after periods of data loss and to reduce the impact on subsequent estimates. The hypothesis that such a threshold existed which would improve orientation was not supported by the results of these experiments. Differing threshold sizes had no impact on the average accuracy of orientation calculated using either approach. Ottosen et al. [404] found in time series air-quality

measurements that linear interpolation across varied gap sizes also had no effect on RMSE. This raises the question why filling only small gaps did not boost the orientation accuracy. It appears that on occasion it actually did, there were many times when it estimated missing data well but averaging across gap filling examples meant that the average RMSE was heavily penalised by occasions where it inaccurately filled data. Examination of the findings (Appendix B section 11.2) showed how well the data were filled was dependent on where in the signal the data loss had occurred and how much data had been lost locally (multiple close occurring gaps) [382, 398]. By analysing the results visually, it became clear that where the data is lost is more important than the gap size. Figures 7.11 and 7.12 show examples of this in action where for example (figure 7.11) the linear and previous perform good gap filling of small gaps, and (figure 7.12) good gap filling of much larger gaps in comparison to the no gap filling method.

(a) X-axis tibia data

(b) Y-axis tibia data
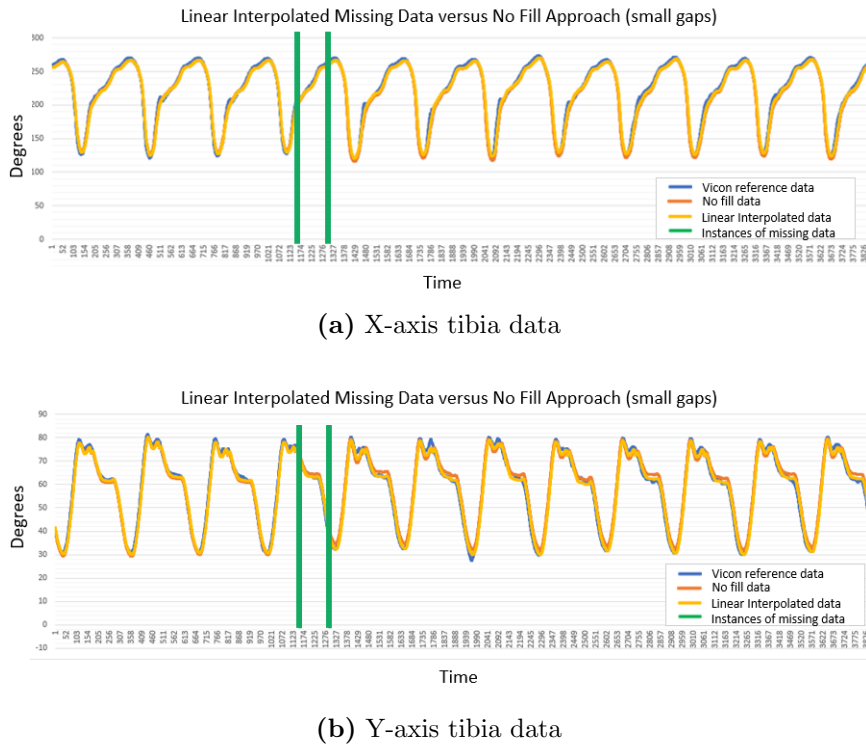
Figure 7.11: Orientation signals stay converged at the start of this section after periods of no data loss. After a small period of data loss (green lines), the filled signal (yellow) accurately fills the missing data better than the no fill data (orange) until it re-converges.

(a) X-axis tibia data

(b) Y-axis tibia data
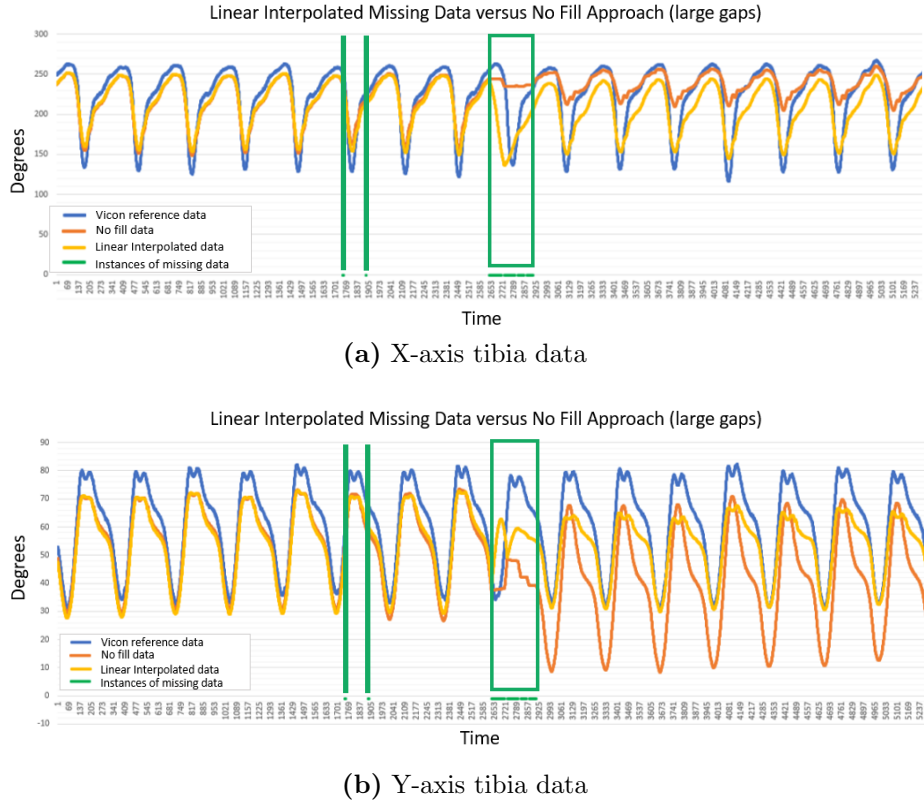
Figure 7.12: Orientation signals stay converged at the start of this section after periods of no data loss. After a large period of data loss (green square), the filled signal (yellow) accurately fills the missing data better than the no fill data (orange) until it re-converges.

In contrast, two more examples show instances when the linear and previous performed bad gap filling of small gaps (figure 7.13), and bad gap filling of much larger gaps (figure 7.14) in comparison to the no gap filling method. The location of the gap is explored more in the Appendix B, where it is shown that the orientation calculated after filling artificially created gaps (using linear interpolation) in slow moving regions of the inertial data, compared to fast moving regions of the inertial data produced different signals (see Section 11.2).

Regarding multiple close-occurring gaps, it can be seen that the Madgwick filter has the ability to converge back to the true orientation after periods of signal interruption due to data loss because of the inherent algorithm design. Figure 7.15 shows the orientation re-convergence which occurred after the data loss seen in figure 7.12. From our experiments it could be seen that generally the algorithm converged back to the correct values quicker when no missing data was filled, than when linear or previous sample filling was used, and the missing values were estimated poorly. It is possible that poorly filled inertial data lead to a poor orientation estimate which was seen to have a negative and cumulative effect on subsequent orientation estimates, which then required more time for the

147
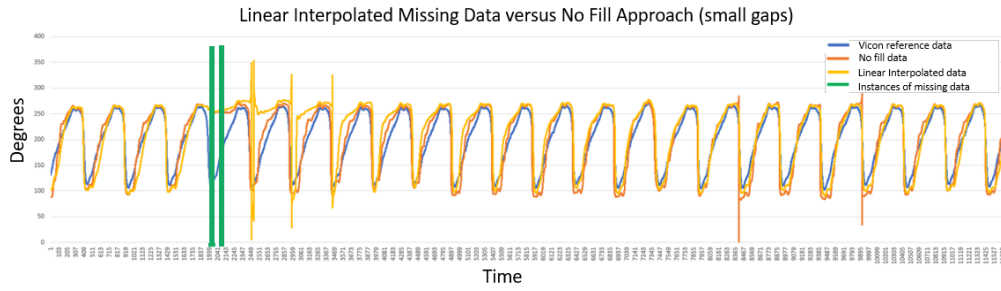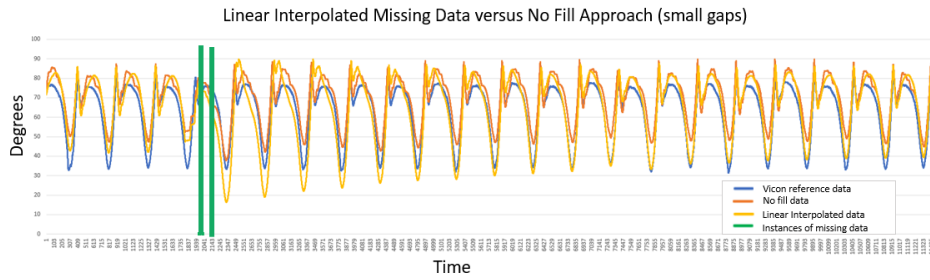
(a) X-axis tibia data

(b) Y-axis tibia data

Figure 7.13: Orientation signals start converged after periods of no data loss. After a small period of data loss (green lines), the filled signal (yellow) fails to accurately fill the missing data versus the no fill data (orange) until it re-converges.
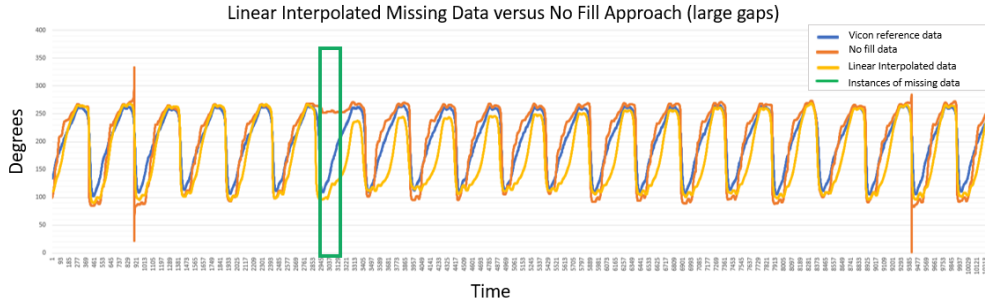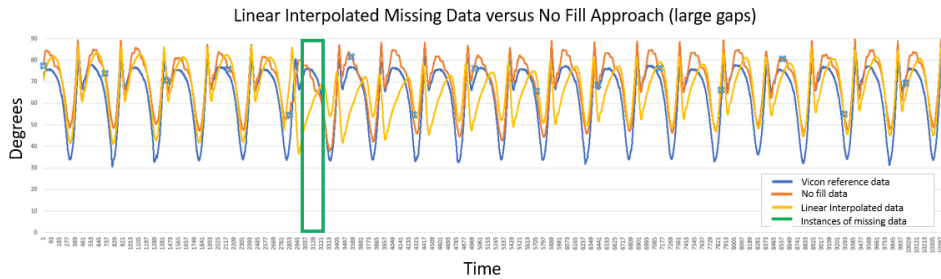
(a) X-axis tibia data

(b) Y-axis tibia data

Figure 7.14: Orientation signals start converged after periods of no data loss. After a large period of data loss (green square), the filled signal (yellow) fails to accurately fill the missing data versus the no fill data (orange) until it re-converges.

algorithm to converge back to the correct values.
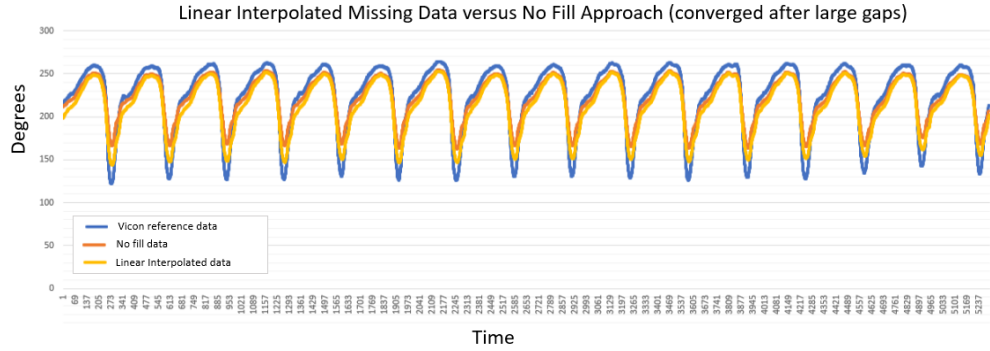
(a) X-axis tibia data
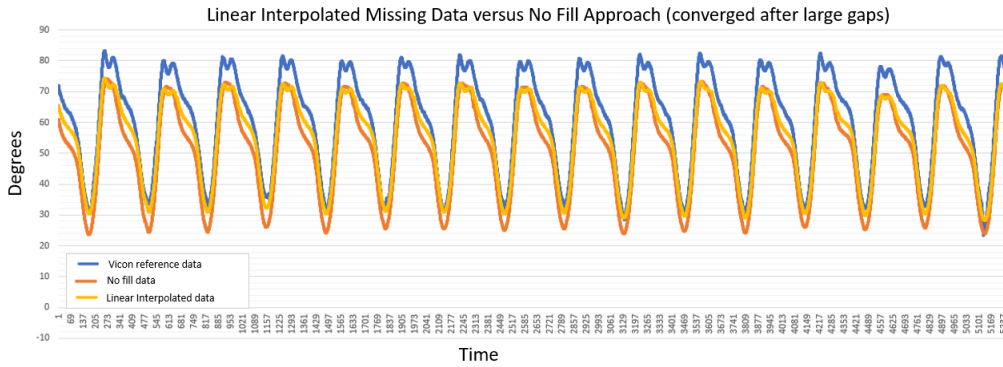
(b) Y-axis tibia data

Figure 7.15: This is an example of the data converging to the correct orientation after the data loss which occurs in figure 7.12.

Figure 7.16 shows an example of this where data were lost before the comparison section with Vicon reference data, and its effects are still visible at the beginning of the plot. As only small amounts of missing data follows closely after, the algorithm is able to converge and correct itself (as seen in the figure). However, If this original data loss was closely followed by significantly more missing samples before the filled orientation had a chance to correct itself, the errors compound and the signal is distorted until it processes enough loss free samples to converge to the correct orientation (see figure 7.17).

Additionally, the results found that data pre-filled with linear interpolation or previous sample filling did not provide significantly more accurate orientation data on average. It is important to remember that the accuracy of linear interpolation is dependent on the edge points of the gap of missing data. The linear interpolation approach draws a straight-line approximation between the last known data point and the next known point, whereas the previous sample filling carries forward the

(a) X-axis tibia data

(b) Y-axis tibia data
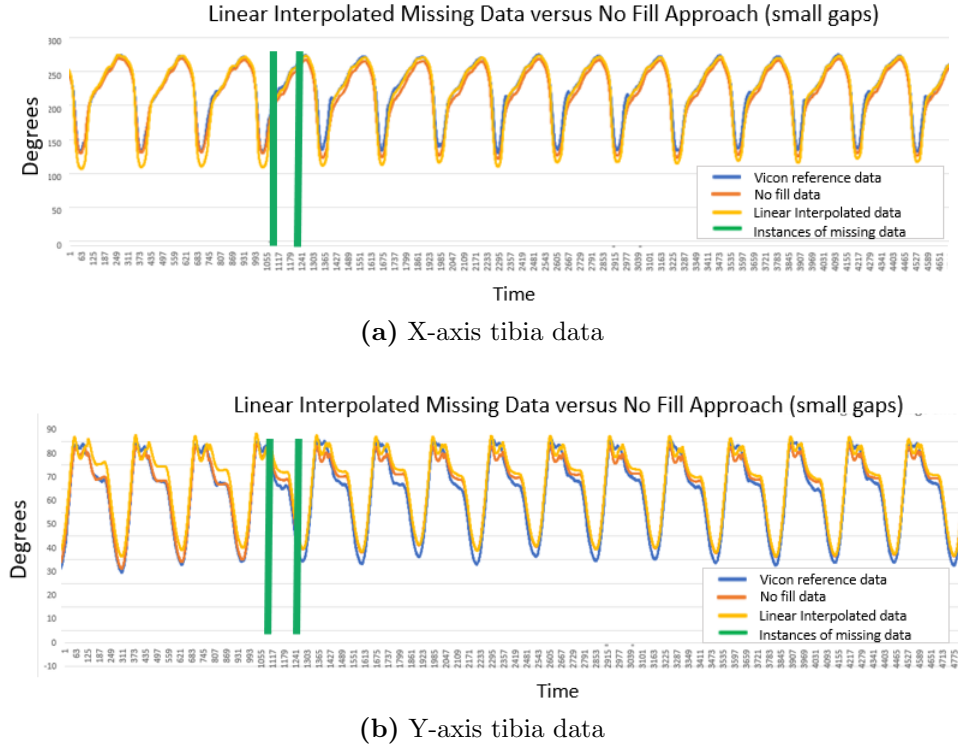
Figure 7.16: Example showing multiple occurrences of data loss. Data loss occurred before this comparison section and it can be seen that the start of the orientation signal is still disturbed by this (yellow). Following this, the algorithm can process enough loss-free data to re-converge, as can be seen where the filled orientation (yellow) re-aligns with the no fill data (orange).

last seen sample before the gap to fill the missing data. When most gap sizes are quite small, as we know is true for this data, and the magnitude of the end points are similar, the difference in linear interpolation across end points versus previous sample filling was not hugely evident. This explains why they performed so similarly in the experiment where many small gaps occurred. This is also true for larger gap sizes when the magnitudes of the end points are close in value, and the missing data may represent a peak that should exist between them. As single imputation methods do not take this variation into account, only endpoints, the values estimated by linear interpolation and previous sample filling approximate similarly. This similarity of performance was also previously found for gap filling heart rate data in Lin et al. [376]. The amount of gap-free orientation data versus gap-affected allowed the algorithm time to converge so that the actual differences in the ability to fill gaps may have been masked averaging across the signal, however, if the interpolation methods had shown major improvement of orientation data, this improvement would be more noticeable comparing performance averages. The aim of trying these approaches to speed up the convergence after data loss was not successful, and was no better on average than not filling the inertial data.

(a) X-axis tibia data

(b) Y-axis tibia data
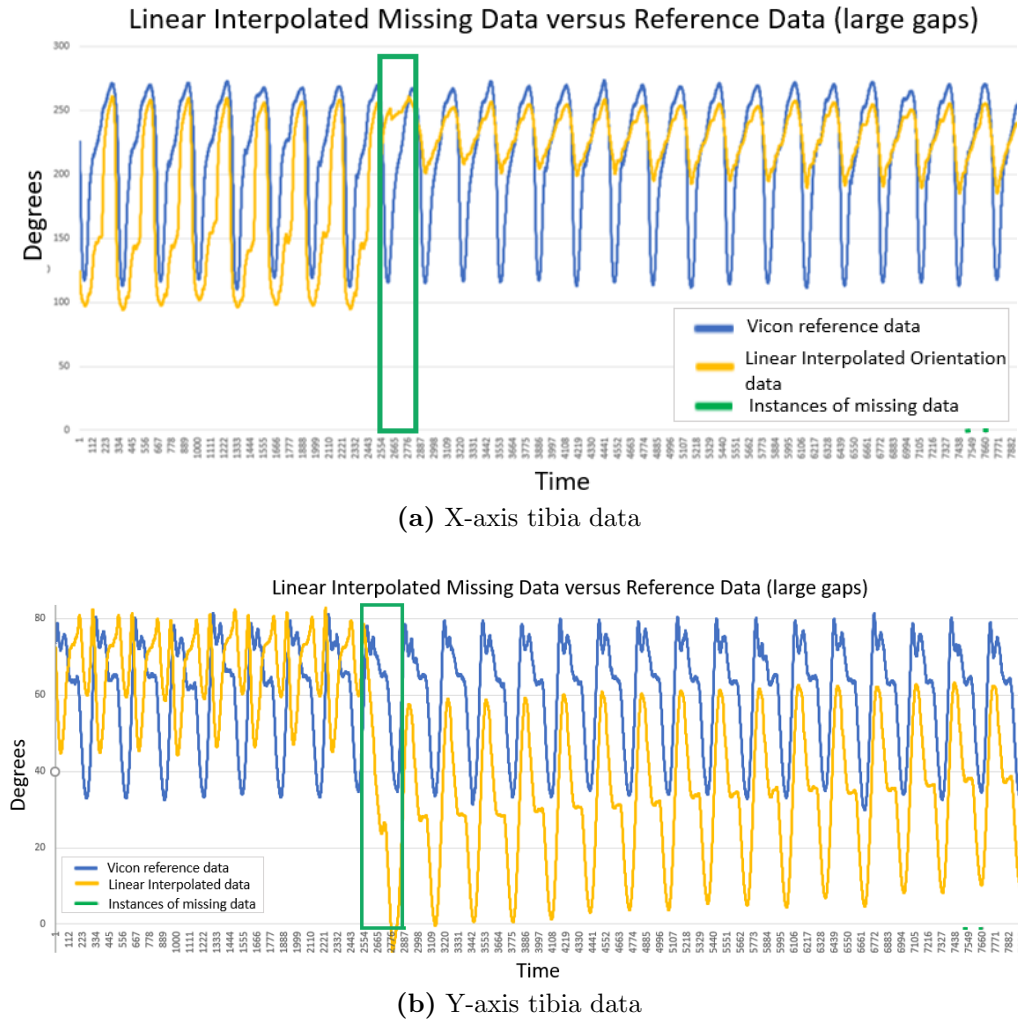
Figure 7.17: Example showing multiple occurrences of data loss, lasting the length of the green section. Data loss occurred before this comparison section and it can be seen that the start of the orientation signal is still disturbed by this (yellow). When further significant data loss occurs (green) before the algorithm can process enough loss-free data to re-converge, the errors grow until re-convergence can take place.

This research found that post-filling also had no effect on the orientation accuracy. There are no studies which appear to have investigated whether the type of post-filling effects orientation accuracy. The post-filling process was performed originally to keep the data aligned with the reference data, and whether this was completed with linear or previous sample filling did not have a major impact on orientation accuracy as seen in the results. This is mainly due to the fact that this step is carried out after the orientation estimates are already calculated using the Madgwick Filter, and so it does not have a knock-on effect on the accuracy of subsequent orientation estimates (unlike pre-filling). Although this did not improve the orientation accuracy, this post-processing step is important to be able to accurately estimate particular events in the data e.g., the orientation of the foot at initial

contact with the ground during running, in which case either linear interpolation or previous sample filling seems appropriate.

Another source of the observed large errors in orientation stem from the use of 6DOF as opposed to 9DOF. Figures 11.14 (a) and (b) show the results of an experiment (included in the Appendix C 11.3) where data was collected on one participant collecting running data. Data were processed with 6DOF and 9DOF and compared to the reference Vicon data collected. The 9DOF data successfully avoids convergence errors compared to the 6DOF calculation after periods of data loss, from which errors begin to accumulate during the high impact running in the 6DOF data. This confirms that the convergence after data loss issue examined within this thesis is unique to 6DOF estimation, as the magnetometer data provides more accurate orientation data without long periods of convergence. It should be noted however, it is possible that large amounts of missing data would still pose an issue for 9DOF estimation [373], but this is beyond the aims of this thesis.

7.4.4   Limitation

The main limitation of this work is that optimising the Madgwick filter parameters was prioritised after addressing missing data [optimising the Madgwick filter is explored in the next sections (Study 1.3A and 1.3B)]. Optimising this parameter improves the errors caused by data loss as the beta parameter affects overall convergence of the algorithm including after disturbances. We tried to improve this convergence through including imputed missing data which was unsuccessful. A clear limitation to this work is not performing this gap filling analysis across a range of sampling rates, to see how that also impacts the results, with obviously lower sampling rates reducing the impact of missing data on orientation.

7.4.5   Future Work

Other methods for reducing the effect of dropped data on orientation should be explored. Future work could investigate the possibility of developing gap filling approaches which are based on using different interpolation methods depending on the content in the signal around the point of data loss, and so a pattern matching scheme could identify the best approach to fill data based on the cyclic nature of these signals. This type of pattern matching was examined with multiple imputation by Azimi et al. [377] on missing heart-rate data and has been used in non-health time series data by Mariethoz et al. [378]. This would establish rules in how to handle data loss for the different events

in the data. Similarly, future work could consider using a machine learning approach to handle the treatment of data loss through training on observed values, although problems can occur with larger gap sizes where predicted values of one gap are used to predict the next [380]. Also, due to the variability of this movement data this can create problems as a model might have to be trained on an individual subject's missing data, rather than broadly a generic gap-filling task [376]. Future work should consider optimising the Madgwick filter parameters prior to addressing missing data, due to the effect that the beta parameter has on the filters overall convergence. None of these approaches are considered in the remainder of this thesis.

7.4.6 Conclusion

The hypothesis that filling some of the inertial data with linear and previous sample filling would result in more accurate orientation estimates was not supported by the results of this experiment. As a whole the data filling approaches (linear and previous) did not have an improvement in accuracy, with linear and previous sample filling performing similarly to not pre-filling the inertial data. While it can be seen above in some situations the linearly pre-filled data did provide more accurate orientation over not filling the data, this increase in accuracy was not seen on average across all the data. As this increase in accuracy may be related to specific ideal situations relating to the content of the signal, where the data loss occurred and the number of local gaps, this would be extremely complex to implement consistently without a model-based approach.

7.5 Algorithm Modification 1: Optimisation of the Beta Parameter (Study 1.3A)

The results in the previous section disproved the hypothesis that using a data pre-processing pipeline to fill missing inertial data would lead to more accurate orientation data. Therefore, no data filling pre-processing was included in Study 1.3 (A or B). Due to the insufficient accuracy of the results for running orientation estimation in Study 1.1 (Section 7.3), further exploration to improve the running orientation accuracy was carried out. Two algorithm modifications to improve the convergence of the orientation estimates will be investigated in Studies 1.3A and 1.3B.

7.5.1 Aim

The aim of this section (Study 1.3A) was to evaluate whether an improvement in our baseline running orientation accuracy could be obtained through modification of the orientation algorithm design, specifically through modification of the beta parameter (β) within the Madgwick filter. A number of studies using both 6DOF [7, 210, 216] and 9DOF [12, 191, 366] have examined the optimisation of the beta parameter within the Madgwick filter, by tuning the fixed value of this parameter to find the optimal beta value which results in the best reduction in orientation errors. This parameter has an important role in the reduction convergence errors within the Madgwick filter. The beta parameter represents the weight given to the accelerometer data in the overall orientation calculation, where β = 0 would calculate orientation using the gyroscope only, and β >0 represents the magnitude of influence from the accelerometer [7]. There is a trade-off in optimising this parameter where low beta is more accurate in fast movements (but will suffer more from drift as the closer to zero this becomes the less gyroscope error is being removed by the accelerometer data) and a high beta is more accurate in slow movements but may be noisy [216]. This parameter controls the success of the convergence of the algorithm during movements.

7.5.2 Methodology

7.5.2.1 Data Processing

The baseline 6DOF Madgwick filter [14] is described in the 'Initial Common Methodology' section of this chapter (Section 7.2.2). The beta parameter within the Madgwick filter is a key component in reducing convergence errors during orientation estimation, as it represents the mean gyroscope measurement error. Depending on its value, the beta parameter dictates how much of the estimated orientation is influenced by the accelerometer readings, where with β = 0 orientation was estimated using solely the gyroscope, and β = 1 the accelerometer largely influences the orientation estimate. In Madgwick's original work they detail a means to estimate the beta value; where $\tilde{\omega}\beta$ represents the estimated mean zero gyroscope measurement error of each axis, beta can be approximated by:

$$\sqrt{\frac{3}{4}}\tilde{\omega}_\beta \tag{7.15}$$

The majority of studies which optimise this parameter do so using a grid search approach [216, 366] which optimises beta specifically for their data or trial and error [7, 12]. Experimental approaches to

automatically (without reference data) estimate sub-optimal parameters have recently been suggested [19], where the value of beta which minimizes the relative orientation difference between two MIMUs aligned and attached to a rigid body is selected. In addition, adaptive beta (not a fixed value) selection procedures have recently been examined [191, 405], where threshold based methods allow the value of beta to be dependent on the inertial sensor vector measurements. This allows beta to be adapted briefly during conditions of high accelerations or high magnetic disturbance in order to reduce errors. However, this is not helpful in running over distances where the presence of high accelerations are constant and there is no variation to slower movements once the running begins, as reflected in the dataset of this thesis. Therefore only the traditional grid search approach for parameter optimisation was considered [12, 366]. This also ensures that the methodology in this section (Study 3A) remains the same in comparison to the baseline in Study 1.1, except for the change in beta parameter value. Beta was set to 0.033 in Study 1.1, following the guidance provided in Madgwick's original work [14].

A grid search approach was carried out to calculate orientation across each axis (X, Y, Z) of the tibia, thigh and sacrum. A subset of the participants was used in the grid search in order to optimise beta on half of the data, and examine its generalisation to the full dataset [366]. As the results in Study 1.2 (Section 7.4) were worsened by the presence of data loss, the most stable 10 participants were chosen for the grid search to remove issues related to outliers. The grid search was carried out with beta values ranging from $\beta = 0.1$ - 0.9 in increments of 0.05, as well as the baseline beta value suggested in Madgwick's original work $\beta = 0.033$, resulting in a set of 14 beta values to be used to calculate orientation data for each segment. For each segment axis (X, Y, Z) the calculated averages for each beta parameter were compared to evaluate the most accurate beta value for each segment (see table 7.19). The best beta across the majority of segments was then applied to the full dataset and compared to the baseline results from Study 1.1.

7.5.2.2 Data Analysis

Normality was assessed by Shapiro-Wilk's test in SPSS to determine whether the normality assumption held for parametric analysis. As data normality was violated, Wilcoxon Signed Ranks tests were performed to assess whether there was a significant difference in population mean ranks of the accuracy between the optimised beta value (lowest average RMSE across segments) and default value used in the baseline algorithm in Study 1.1 (β=0.033). The level of significance was set at $p < 0.05$. The statistical analysis was completed using IBM SPSS Statistics 24.

7.5.3 Results

The grid search analysed the effect of the beta parameter on orientation estimation using the Madgwick 6DOF filter, these results can be seen in table 7.19. It can be seen from the results that for the tibia, thigh and sacrum data, a beta value of 0.10 resulted in the lowest orientation error for each of these segments. $\beta = 0.10$ was selected as the best optimised beta to apply across all segments to the rest of the data.

Table 7.19: Grid search analysis performed on the 10 most stable participant data from Study 1.1. The results show the magnitude of average orientation accuracy in degrees RMSE for each beta value across each axis (X, Y, Z) of the tibia, thigh and sacrum.

RMSE (Degrees)	0.033	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50	0.60	0.70	0.80	0.90
Tibia X	18.01	12.23	12.14	12.23	12.40	12.63	12.89	13.18	13.49	14.17	14.91	15.68	16.49	17.03
Tibia Y	7.37	5.54	5.68	5.89	6.13	6.39	6.66	6.95	7.24	7.86	8.50	9.17	9.87	10.51
Tibia Z	11.82	9.40	9.53	9.81	10.11	10.37	10.80	11.50	11.69	12.45	13.04	14.04	18.17	20.48
Average:	12.40	9.05	9.12	9.31	9.55	9.79	10.12	10.54	10.81	11.49	12.15	12.96	14.84	16.01

RMSE (Degrees)	0.033	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50	0.60	0.70	0.80	0.90
Thigh X	18.72	17.32	17.94	18.66	19.46	20.31	21.20	22.12	23.07	25.03	27.06	29.15	31.30	32.66
Thigh Y	4.69	3.85	3.89	3.95	4.04	4.14	4.25	4.37	4.49	4.75	5.03	5.31	5.60	5.93
Thigh Z	11.45	11.21	11.92	12.56	13.31	14.40	15.61	16.13	16.89	18.57	20.26	22.31	24.96	25.70
Average:	11.62	10.8	11.25	11.72	12.27	12.95	13.69	14.21	14.82	16.11	17.45	18.92	20.62	21.43

RMSE (Degrees)	0.033	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50	0.60	0.70	0.80	0.90
Sacrum X	10.71	8.61	9.13	9.76	10.42	11.09	11.77	12.44	13.10	14.43	15.76	17.08	18.41	19.10
Sacrum Y	6.53	6.67	6.73	6.78	6.83	6.88	6.91	6.95	6.98	7.03	7.07	7.09	7.10	7.18
Sacrum Z	6.13	6.63	7.37	8.11	8.89	9.65	10.40	11.05	11.76	13.11	14.51	15.92	17.48	18.82
Average:	7.79	7.31	7.75	8.22	8.71	9.21	9.69	10.15	10.62	11.52	12.44	13.36	14.33	15.03

The optimised beta ($\beta = 0.10$) was applied to the full data. This analysis was completed twice, once with the full dataset, and again with the full dataset minus apparent outliers (see figure 7.18). It is difficult to define whether bad data points were in fact outliers. These were results with high error across various beta values linked to the occurrence of data loss but also due to irregularities which occurred in the signal simply due to 6DOF tracking (See Appendix C). Both of these issues lead to poor algorithm convergence which could not be improved by beta tuning. Outliers were defined as these samples where the results did not stabilise across any varied beta value.

There were samples which were highly inaccurate for $\beta = 0.033$ but as β increased, the errors reduced, and the signal became stable. These were not considered as outliers but simply a symptom of an untuned beta parameter, as improvement from parameter tuning could be easily seen. For this reason, these data were included in the main analysis (despite inflating the average error).

Running segment orientation errors were reduced by 6.4, 1.5, and 2.0 degrees on average across the (full dataset) tibia, thigh and sacrum, respectively. Running segment orientation errors were reduced by 3.8, 1.7, and 1.3 degrees on average across the (outlier removed) tibia, thigh and sacrum, respectively. It can be seen that the average magnitude of error reduction was similar across both

datasets axes (full and outliers removed).

Full Dataset

RMSE (Degrees)	β= 0.033	β= 0.10
Tibia X	26.2	15.1
Tibia Y	10.5	6.6
Tibia Z	15.2	11.0
Average:	17.3	10.9

RMSE (Degrees)	β= 0.033	β= 0.10
Thigh X	19.2	16.4
Thigh Y	4.8	3.8
Thigh Z	11.8	10.4
Average:	12.0	10.2

RMSE (Degrees)	β= 0.033	β= 0.10
Sacrum X	22.7	16.8
Sacrum Y	7.0	7.0
Sacrum Z	17.2	17.1
Average:	15.6	13.6

Outliers Removed

RMSE (Degrees)	β= 0.033	β= 0.10
Tibia X	20.7	12.2
Tibia Y	8.1	5.6
Tibia Z	12.6	9.2
Average:	13.8	9.0

RMSE (Degrees)	β= 0.033	β= 0.10
Thigh X	18.6	16.2
Thigh Y	4.8	3.8
Thigh Z	11.3	10.3
Average:	11.6	10.1

RMSE (Degrees)	β= 0.033	β= 0.10
Sacrum X	15.1	11.5
Sacrum Y	6.5	6.5
Sacrum Z	8.9	8.6
Average:	10.2	8.9

Figure 7.18: This figure shows a comparison of the results from Study 1.3A and Study 1.1. In the current study, the optimal beta value of β = 0.1 was applied to the entire dataset, and in Study 1.1 the default beta value of β = 0.033 was applied to the entire dataset. This comparison was also made for a reduced dataset where outliers were removed. Overall, it can be seen that the tuning of the beta parameter provided a reduction in orientation error for both datasets.

Wilcoxon Signed Ranks tests on Tibia and Thigh (X, Y and Z) data consistently reported that there was a significant reduction in orientation error using the optimised beta parameter in comparison to the baseline beta (for both the full dataset and outlier removed dataset). Sacrum data (using full dataset) found no significant relationship in axes X or Y, but a statistically significant reduction in error due to beta tuning on the Z-axis. The results of the non-parametric tests can be seen in table 7.20.

7.5.4 Discussion

The primary aim of this section was to optimise the beta parameter associated with the Madgwick filter and quantify the resulting change in orientation accuracy compared to the baseline orientation accuracy from Study 1.1, in an attempt to reduce running orientation errors. The findings support the original hypothesis that the modified implementations of the baseline algorithm with parameter tuning would provide an improvement in accuracy over the baseline implementation without modifications.

Table 7.20: Wilcoxon Signed-Rank tests comparing average accuracy across all participants (n=20) of optimised and baseline beta values across each axis (X, Y, Z) of the tibia, thigh and sacrum.

Segment Axis	Full Dataset		Outliers Removed	
	Z	p	Z	p
Tibia-$X_{0.033}$ - Tibia-$X_{0.1}$	-10.1	5.0E-24*	-8.9	4.7E-19*
Tibia-$Y_{0.033}$ - Tibia-$Y_{0.1}$	-8.7	4.0E-18*	-7.3	3.5E-13*
Tibia-$Z_{0.033}$ - Tibia-$Z_{0.1}$	-8.2	7.2E-17*	-7.1	9.5E-13*
Thigh-$X_{0.033}$ - Thigh-$X_{0.1}$	-3.4	1.0E-3*	-3.1	2.0E-3*
Thigh-$Y_{0.033}$ - Thigh-$Y_{0.1}$	-5.0	7.3E-7*	-5.1	3.6E-7*
Thigh-$Z_{0.033}$ - Thigh-$Z_{0.1}$	-2.6	0.01*	-2.3	0.02*
Sacrum-$X_{0.033}$ - Sacrum-$X_{0.1}$	-1.6	0.11	-0.8	0.45
Sacrum-$Y_{0.033}$ - Sacrum-$Y_{0.1}$	-1.5	0.15	-1.9	0.53
Sacrum-$Z_{0.033}$ - Sacrum-$Z_{0.1}$	-2.3	0.02*	-3.7	2.0E-4*

Note: * denotes statistical significance.

The improvement in orientation error due to the beta parameter tuning was statistically significant for all axes of the tibia and thigh data for both the full dataset, and the dataset with outliers removed. However with regards to the sacrum, although the parameter optimisation improved the average orientation accuracy across axes (X, Y, Z), and performed better than or equivalent to the baseline beta parameter on an axis level, there were conflicting results on the statistical significance of optimising the parameter when comparing sacrum axis-level performance, where the both X and Y axes results pointed to general non-significance (see table 7.20). This can be explained by the results contained within table 7.19, where it can be seen that the best 10 participant sacrum data had a small magnitude of improvement (error reduction due to β parameter tuning) of the average results between the two beta values (0.033 and 0.1) generally, and when compared to the tibia specifically. Although the thigh segment has similar magnitudes of variation in averages to the sacrum, it can be seen that on an axis-level the parameter tuning lead to improvements in accuracy on each thigh axis, whereas parameter tuning only lead to improvements on one sacrum axis (X-axis). This is shown in table 7.19, where analysis was performed on the 10 best participant data only, but this is reflective of the results across the full dataset (20 participants X 9 comparison sections between inertial data and Vicon = 180 independent sample results on the algorithms performance) where sometimes where orientation error was already low, parameter tuning lead to increased orientation error. The lower level of error may be in part due to lower levels of impact accelerations on the sacrum (compared to lower body segments) causing the inertial data and resulting orientation to suffer from less error due to noise. It was clear that the parameter tuning performed best and reduced error the most on noisy data. Therefore, it is possible due to the generally lower error associated with the sacrum (see outliers removed) parameter

tuning was less effective. Although parameter tuning was not statistically significant overall for the sacrum segment, on average parameter tuning lead to reductions in error. The inflated average error of the sacrum in figure 7.18 compared to the best 10 participants 7.19 was due to the inclusion of a number of (valid non-outlier) samples which performed very poorly on lower beta values ($\beta = 0.033$) but improved with parameter tuning.

The value 0.10 was shown to produce the lowest orientation errors for the stable dataset (n=10). Optimal beta values were found with respect to individual segment, and the majority best beta value was used to calculate optimised parameter orientation accuracy results for the full dataset (n=20). The optimal beta parameter of 0.10 was shown to generalise well for the rest of our dataset but performed better on the tibia and thigh data than the sacrum. Although the stable participant dataset (n=10 participants) was used in the grid search to optimise the beta parameter, the most stable data (while free from outliers and valid non-outlier unstable data) still had rather large errors before beta parameter tuning. See figure 7.19 where the baseline results for the most stable 10 participants had an average error across axes of 12.4, 11.6 and 7.8 degrees across the tibia, thigh and sacrum, respectively. This is due to data loss discussed in the previous section. Data loss can cause the 6DOF algorithm to deviate from the correct orientation estimates, until the algorithm (with the help of the beta parameter) converges back to the correct orientation estimate after some time. As all segment data orientation had to be calculated sequentially, periods of data loss which occurred just before/or during the segmented data (sections to compare between sensor data and Vicon) could greatly affect the resulting orientation estimates during that 1-minute data comparison. This means that the cumulative errors caused by data loss still affected the subsequent orientation estimates and could not always be avoided. This occurred for convergence errors caused by data loss and convergence errors specific to 6DOF.

In general, low values of beta ($\beta <=0.20$) tended to result in better orientation estimates (see table 7.19), where a low beta value reflects that the majority of the orientation estimate was controlled by the gyroscope orientation estimate. This is most likely due to the evaluation of running movements, during which the gyroscope is better at estimating orientation than the accelerometer. The accelerometer performs well during slow or static periods which were not included in this experimental design. It is likely the high beta values were greater disturbed by external acceleration and high impacts felt by the accelerometer during running.

There are many studies which also perform beta parameter tuning to improve their results [7,

12, 19, 191, 210, 216, 366]. Optimal beta values were very much dependent on the data and specific experimental design of each study.

7.5.5 Future Work

Further exploration of beta parameter optimisation could be investigated with respect to a varying, rather than fixed value [191, 405]. This is the approach used in Kalman filtering which allows the filter gain (beta) to change with respect to changes in movement or speed. This has been explored in a number of studies which use a complementary filter [12] but not with respect to running orientation estimation. It is possible this could further reduce the orientation errors due to signal drift and large impact accelerations during running. This idea will not be explored in the current thesis.

7.5.6 Conclusion

This section implemented and evaluated the effect of beta parameter optimisation on running orientation estimation errors at the tibia, thigh and sacrum. A beta value of 0.1 was found to generalise well to the full dataset and provided a significant reduction in error for the tibia and thigh segments. This is a common approach undertaken in 9DOF orientation estimation research [12, 19] and can also be seen in 6DOF orientation estimation [366], but not with respect to running.

7.6 Algorithm Modification 2: Optimisation of Stochastic Gradient Descent (Study 1.3B)

7.6.1 Aim

As the results for the running segment orientation accuracy in Study 1.1 were below what would be considered acceptable for human movement analysis (>5 degrees error [367,368]), studies 1.3 A and B examined whether a modification of the Madgwick algorithm design could improve the 6DOF running orientation estimation to acceptable levels of accuracy. In this section (Study 1.3B) a modified version of the 6DOF Madgwick Filter [14] was examined by investigating the optimisation steps involved in the algorithm. In order to improve the convergence of the Madgwick filter, the performance of a different Stochastic Gradient Descent optimiser was assessed in place of the traditional stochastic gradient descent code block in the algorithm design. Other Stochastic Gradient Descent optimisation schemes exist such as: momentum, Adagrad, Adam, which are commonly used in the domain of artificial intelligence and deep learning. It is possible that using a different stochastic gradient descent optimiser e.g., momentum [406], may improve running orientation estimates, no studies have investigated this to-date.

7.6.2 Methodology

7.6.2.1 Data Processing

The baseline 6DOF Madgwick filter [14] is described in the 'Initial Common Methodology' section of this chapter (Section 7.2.2). The Madgwick filter is designed to first calculate an orientation estimate using the gyroscope signal alone through integration. As gyroscope integration cannot exclusively be used to calculate orientation due to accumulating signal drift (see Section 3.2), the accelerometer signal is used to offset the error in the gyroscope based orientation estimate.

Modifications were implemented with respect to the Gradient Descent element of the algorithm. Unlike in traditional Stochastic Gradient Descent where one sample (sensor measurements at time t) is iterated over a number of times to traverse the error surface for each measurement, in Madgwick's filter there is only one iteration of Stochastic Gradient Descent per sample at time t, in order to reduce the computational load [14]. In preliminary experiments of this thesis, the introduction of multiple iterations of SGD were examined, which lead to small improvements in accuracy. The preliminary

experiments involved creating new parameters for the Stochastic Gradient Descent algorithm, numbers of iterations and different step sizes. Algorithm performance during a real-time running application remained an important factor, with this in mind, few iterations (less computations) and smaller step sizes (resulting in no additional noise added), were investigated and a general approach was designed.

In addition to this there are a number of known optimisers for SGD (Momentum, Nesterov accelerated gradient, Adagrad, Adam [369]) which are used in other research areas relating to Machine Learning and Deep Learning for solving optimisation problems [407, 408]. No studies have investigated the performance of optimisers other than SGD embedded in the popular Madgwick Filter. In this thesis, building on from the intermediate stage of multiple iterations from the preliminary experiments, momentum with multiple iterations of Stochastic Gradient Descent was explored.

A modified algorithm was implemented by replacing the Stochastic Gradient Descent component with multiple iterations of Momentum based Stochastic Gradient Descent. Different parameters for the number of iterations (2, 4, 6), step size (0.01, 0.02, 0.06) and beta parameter (0.033 and 0.1) were examined within the multiple iteration framework in combination with various values for the momentum parameter (0.3, 0.5, 0.7, 0.9). Momentum allows the optimisation process central to SGD to converge quicker, so it was hypothesised that this may lead to reducing the convergence errors seen in Study 1.1 [408, 409]. This analysis was carried out as a grid search, initially on the best 10 participant subset as mentioned in Study 1.3A.

A grid search was set up with each combination of: number of iterations, step size, momentum parameter and beta, to identify the parameter value combination with the lowest average RMSE across segment axes for each of the two explored beta values (baseline β=0.033 [Study 1.1] and optimised β=0.1 [Study 1.3A]). Accelerometer and gyroscope data were used to calculate orientation using these parameters in the modified Madgwick filter. Orientation estimates were computed for all segments and separated into comparison sections to compare with the Vicon reference data using RMSE (following methodology as in Study 1.1). The missing samples in the inertial data were not filled using a pre-processing data filling pipeline, as the results in Study 1.2 (Section 7.4) showed that the explored methods did not improve orientation accuracy. However, linear gap filling was performed to fill any missing data after orientation data were calculated in order to keep the sensor data and Vicon data time-aligned. Once the optimal parameter combinations (lowest RMSE) were identified for each segment using the grid search, these were applied to the full dataset (n=20) and the 'outlier

removed dataset' as explained in Study 1.3A.

7.6.2.2 Data Analysis

RMSE was used to compare the orientation estimates from the inertial data and the Vicon reference data. The level of accuracy of this approach was assessed and compared to the baseline accuracy in Study 1.1 (β=0.033), and the optimised beta from Study 1.3A (β=0.1). Normality was assessed by Shapiro-Wilk's test in SPSS to determine whether the normality assumption held for parametric analysis. As data normality was violated, Wilcoxon Signed Ranks tests were performed to assess whether there was a significant difference in accuracy between the baseline orientation algorithm in Study 1.1 (β=0.033) and the modified orientation filter described in this section. The level of significance was set at $p < 0.05$. The statistical analysis was completed using IBM SPSS Statistics 24.

7.6.3 Results

The analysis of the results were performed at a segment axis level. The grid search analysis of the best 10 participant dataset for each segment showed that there were different optimal parameter combinations for each beta parameter (baseline = 0.033 or optimised=0.1), and also for each segment. The best parameter combination using the baseline beta will be referred to as BestParam0.033 and the best parameter combination using the optimised beta will be referred to as BestParam0.1. Tables 7.21, 7.22 and 7.23 show the resulting RMSE orientation accuracy for each segment when these best parameter combinations were applied to the full dataset and outlier dataset.

Wilcoxon Signed Ranks tests were performed on each segment axis (X, Y, Z) to investigate whether the changes in orientation error between the benchmark accuracies (baseline and optimised beta) and the best parameter sets (BestParam0.033 and BestParam0.1) were significant. Although there were a number of significant relationships, when using the means to identify the direction of the relationship, the general trend across segments showed that for the majority of axes pairs, the best parameter sets (BestParam0.033 and BestParam0.1) were not significantly more accurate than the baseline beta = 0.033, and optimised beta = 0.1 calculations.

Statistical Analysis Results of the Tibia Segment data

The results (see table 7.21) show that the parameter combination of: 6 iterations, 0.01 learning rate, and 0.3 momentum, optimised the tibia data with respect to $\beta = 0.033$ (baseline), and parameter

combination of: 2 iterations, 0.01 learning rate, and 0.5 momentum value optimised the tibia data with respect to β = 0.1 (optimised from Study 1.3A).

Table 7.21: This figure shows the results of the segment running orientation accuracy across the tibia on the full dataset. Orientation was compared to the orientation accuracy of the results from studies 1.1 (baseline beta β = 0.033) and 1.3A (optimised beta β = 0.1). The top performing set of parameters for each beta value, 0.033 and 0.1, are referenced by BestParam0.033 and BestParam0.1 in the figure, respectively.

Full Dataset

RMSE (Degrees)	β= 0.033	BestParam0.033	β= 0.10	BestParam0.1
Tibia X	26.2	25.9	15.1	15.1
Tibia Y	10.5	10.2	6.6	6.7
Tibia Z	15.2	14.7	11.0	11.2
Average:	17.3	16.9	10.9	11.0

Outliers Removed

RMSE (Degrees)	β= 0.033	BestParam0.033	β= 0.10	BestParam0.1
Tibia X	20.7	21.0	12.2	12.2
Tibia Y	8.1	8.1	5.6	5.7
Tibia Z	12.6	12.0	9.2	9.5
Average:	13.8	13.7	9.0	9.1

Statistical Analysis Results of the Thigh Segment data

The results (see table 7.22) show that the parameter combination of: 6 iterations, 0.01 learning rate, and 0.3 momentum, optimised the thigh data with respect to β = 0.033 (baseline), and no parameter combination further improved the results of the thigh data compared to β = 0.1 (optimised from Study 1.3A).

Statistical Analysis Results of the Sacrum Segment data

The results (see table 7.23) show that the parameter combination of: 4 iterations, 0.01 learning rate, and 0.5 momentum, optimised the sacrum data with respect to beta = 0.033 (baseline), and parameter combination of: 4 iterations, 0.01 learning rate, and 0.9 momentum value optimised the tibia data with respect to beta = 0.1 (optimised from Study 1.3A).

7.6.4   Discussion

The primary aim of this section was to assess whether the modified algorithm improved the segment orientation estimation over the baseline or the optimised beta parameter approaches. The hypothesis was that the modified implementation of the baseline algorithm would provide an improvement in

Table 7.22: This figure shows the results of the segment running orientation accuracy across the thigh on the outlier removed dataset (as explained in Study 1.3A). Orientation was compared to the orientation accuracy of the results from studies 1.1 (baseline beta = 0.033) and 1.3A (optimised beta = 0.1). The top performing set of parameters for each beta value, 0.033 and 0.1, are referenced by BestParam0.033 and BestParam0.1 in the figure, respectively.

Full Dataset

RMSE (Degrees)	β= 0.033	BestParam0.033	β= 0.10	BestParam0.1
Thigh X	19.2	18.7	16.4	16.4
Thigh Y	4.8	5.0	3.8	3.8
Thigh Z	11.8	12.0	10.4	10.4
Average:	12.0	11.9	10.2	10.2

Outliers Removed

RMSE (Degrees)	β= 0.033	BestParam0.033	β= 0.10	BestParam0.1
Thigh X	18.6	18.2	16.2	16.2
Thigh Y	4.8	5.0	3.8	3.8
Thigh Z	11.3	11.5	10.3	10.3
Average:	11.6	11.6	10.1	10.1

accuracy over the baseline implementation. The experimental results fully support the hypothesis that using the optimised algorithm lowers the overall error associated with the running data orientation estimation when compared to the baseline results in Study 1.1.

It can generally be seen that both optimised parameter sets from this section, BestParam0.033 and BestParam0.1, resulted in lower orientation errors compared to the baseline results using β=0.033. However when comparing across sets with the same beta parameter values (β=0.033 versus BestParam0.033 and β=0.1 versus BestParam0.1) the improvements aren't as clear. It can be

Table 7.23: This figure shows the results of the segment running orientation accuracy across the sacrum on the outlier removed dataset (as explained in Study 1.3A). Orientation was compared to the orientation accuracy of the results from studies 1.1 (baseline beta=0.033) and 1.3A (optimised beta=0.1). The top performing set of parameters for each beta value, 0.033 and 0.1, are referenced by BestParam0.033 and BestParam0.1 in the figure, respectively.

Full Dataset

RMSE (Degrees)	β= 0.033	BestParam0.033	β= 0.10	BestParam0.1
Sacrum X	22.7	23.2	16.8	11.8
Sacrum Y	7.0	6.7	7.0	4.9
Sacrum Z	17.2	17.1	17.1	13.8
Average:	15.6	15.3	13.6	10.2

Outliers Removed

RMSE (Degrees)	β= 0.033	BestParam0.033	β= 0.10	BestParam0.1
Sacrum X	15.1	15.8	11.5	9.6
Sacrum Y	6.5	6.2	6.5	4.7
Sacrum Z	8.9	9.2	8.6	9.1
Average:	10.2	10.4	8.9	7.8

Table 7.24: Wilcoxon signed-rank tests comparing average accuracy across all participants (n=20) of baseline beta (0.033) and BestParam0.033 values across each axis (X, Y, Z) of the tibia, thigh and sacrum.

Segment Axis	Full Dataset		Outliers Removed	
	Z	p	Z	p
Tibia-$X_{0.033}$ - Tibia-$X_{BestParam0.033}$	-0.3	0.7	-2.2	0.03*
Tibia-$Y_{0.033}$ - Tibia-$Y_{BestParam0.033}$	-1.8	0.07	-0.9	0.9
Tibia-$Z_{0.033}$ - Tibia-$Z_{BestParam0.033}$	-4.5	<0.01*	-3.7	<0.01*
Thigh-$X_{0.033}$ - Thigh-$X_{BestParam0.033}$	-1.5	0.1	-1.1	0.3
Thigh-$Y_{0.033}$ - Thigh-$Y_{BestParam0.033}$	-2.8	0.01*	-2.8	0.01*
Thigh-$Z_{0.033}$ - Thigh-$Z_{BestParam0.033}$	-1.8	0.08	-2.2	0.03*
Sacrum-$X_{0.033}$ - Sacrum-$X_{BestParam0.033}$	-1.3	0.2	-1.2	0.2
Sacrum-$Y_{0.033}$ - Sacrum-$Y_{BestParam0.033}$	-3.8	<0.01*	-3.6	<0.01*
Sacrum-$Z_{0.033}$ - Sacrum-$Z_{BestParam0.033}$	-1.1	0.3	-1.4	0.2

Note: * denotes statistical significance.

Table 7.25: Wilcoxon signed-rank tests comparing average accuracy across all participants (n=20) of optimised baseline beta (0.1) and BestParam0.1 values across each axis (X, Y, Z) of the tibia, thigh and sacrum.

Segment Axis	Full Dataset		Outliers Removed	
	Z	p	Z	p
Tibia-$X_{0.1}$ - Tibia-$X_{BestParam0.1}$	-1.1	0.3	-2.3	0.02*
Tibia-$Y_{0.1}$ - Tibia-$Y_{BestParam0.1}$	-1.2	0.2	-3.9	<0.01*
Tibia-$Z_{0.1}$ - Tibia-$Z_{BestParam0.1}$	-5.5	<0.01*	-4.3	<0.01*
Thigh-$X_{0.1}$ - Thigh-$X_{BestParam0.1}$	0	1.0	0	1.0
Thigh-$Y_{0.1}$ - Thigh-$Y_{BestParam0.1}$	0	1.0	0	1.0
Thigh-$Z_{0.1}$ - Thigh-$Z_{BestParam0.1}$	0	1.0	0	1.0
Sacrum-$X_{0.1}$ - Sacrum-$X_{BestParam0.1}$	-8.8	<0.01*	-6.2	<0.01*
Sacrum-$Y_{0.1}$ - Sacrum-$Y_{BestParam0.1}$	-8.8	<0.01*	-7.8	<0.01*
Sacrum-$Z_{0.1}$ - Sacrum-$Z_{BestParam0.1}$	-6.07	<0.01*	-5.3	<0.01*

Note: * denotes statistical significance.

seen that the BestParam0.033 parameter set generally improves on or is similar to β=0.033 (with the exception of sacrum outliers removed data). With respect to BestParam0.1, as the grid search found no parameter set was better than β=0.1 for the thigh data, it is mixed results on whether BestParam0.1 provides an improvement over β=0.1. This is due to the fact that it was slightly less accurate for the tibia datasets and improved accuracy on the sacrum dataset.

Comparing β=0.033 versus β=0.1 versus BestParam0.1 for the sacrum, it can be seen this modified approach (BestParam0) lead to significantly lower errors. It is unclear why the BestParam0.1 had such an improvement on the sacrum specifically, but this could be due to the lower impact accelerations experienced by the sacrum segment allowing a clearer estimation of the resulting orientation after the multiple iterations and momentum. The increase in orientation accuracy associated with the modified

approach must be due to the two factors which differ in design from the original approach: multiple iterations and momentum. With respect to multiple iterations, in Madgwick's original algorithm there is only one iteration of Stochastic Gradient Descent per sample at time t, in order to reduce the computational load. Traditionally, stochastic gradient descent iteratively optimises each sample, to reach the minimum point of the error curve (in this case representing orientation error in the accelerometer estimate). It does this by taking multiple steps towards the minimum point by checking the gradient at each step, and taking the next step in the direction of the negative gradient. It is possible that the one iteration SGD is sufficient to remove some of the orientation error, but multiple small steps of SGD (more iterations) may lead to a better error estimate. With respect to momentum, many optimisers exist for stochastic gradient descent which is commonly used in Machine Learning research. Momentum (used in this thesis) is one of the simpler optimisers and helps the SGD to converge faster over multiple iterations, with less oscillation in step size. The hypothesis was that the better convergence may help to reduce orientation errors [408, 409]. One limitation of study 1.3B is that multiple iterations without momentum was not extensively investigated on the whole dataset, and so it is difficult to know whether the combination of iterations and momentum or iterations alone increased the orientation accuracy, but by examining the results it would seem that the majority of the improvement comes from the beta parameter itself. This is suggested by the fact no better parameter combination was found in the grid-search for the thigh data and that differences between tibia β=0.1 and BestParam0.1 were very small in magnitude and additionally, it can be seen that always lower errors resulted from one of the set ups with the improved beta parameter (β=0.1 and BestParam0.1).

Although in general it can be seen that on average across the axes the new sets of parameters resulted in lower orientation errors or similar to the baseline/optimsed comparison, for the majority of the Wilcoxon rank pairs it was found that these changes in accuracy from using the best set of parameters did not have significantly lower error. The most important comparison is between β=0.1 and BestParam0.1, where β=0.1 resulted in the lowest errors in Study 1.3A. Using the best parameter set (BestParam0.1) performed well on the sacrum data but did not universally improve errors across segments. Therefore, this suggests that out of the two modified approaches from Study 1.3A and 1.3B (and without trying different optimisers e.g., ADAM [369]), the simple beta parameter tuning (β=0.1 from Study 1.3A) which has less computation is the recommended approach for parameter tuning.

No other studies in this domain appear to have performed this type of modification to the Madgwick filter which makes it difficult to compare results. Although the results across segments were an

improvement over the baseline results, the magnitude of the orientation errors across segments axes remain too high to be used for absolute estimation of human movement (approximately >10 degrees on most axes) when compared to other studies involving estimation of segment or joint angles (errors of <6 degrees in both [8, 211]).

7.6.5 Future Work

This section explored the modification of the Gradient Descent block within the Madgwick filter by applying an additional optimiser (e.g., Momentum) to the filter (in place of SGD) to evaluate its effect on orientation estimation. Although the orientation errors were not sufficiently reduced to clinically appropriate levels, the results have shown that this approach was successful in reducing some of the error compared to the baseline analysis (Study 1.1). Future work could investigate how this adaptation would perform embedded in a 9DOF orientation estimation algorithm, which would not suffer from the same error experienced with 6DOF. In addition, it is possible that other optimisations of Stochastic Gradient Decent (RMSprop, ADAM, etc.) may allow for better convergence and therefore further reduced orientation errors. This will not be investigated in the rest of this thesis.

7.6.6 Conclusion

No studies appear to have explored the use of Stochastic Gradient Descent optimisers with regard to orientation estimation in the Madgwick filter. In addition, no studies appear to have investigated the use of multiple iterations of SGD within the Madgwick filter. In this section, a modified implementation of the Madgwick filter is examined which combines multiple iterations of SGD, in combination with Momentum, an SGD optimisation method. The results of this analysis showed that orientation estimated with this modified algorithm outperformed the baseline results from Study 1.1 (β=0.033) using both optimised parameter sets BestParam0.03 and BestParam0.1. However, when comparing specifically the baseline versus BestParam0.03 and optimised beta versus BestParam0.1, there was not as clear improvement from using BestParam0.1 over the optimised beta (with the exception of the sacrum). Despite the clear improvements made over the baseline Study 1.1 results, on average the optimised beta best parameters set (BestParam0.1) utilising momentum and multiple iterations did not provide great improvements over simple beta optimisation (β=0.1) alone.

7.7 Overall Summary

This chapter can be broadly summarised as a set of experiments investigating whether the 6DOF orientation errors could be reduced to an acceptable level, in order to be used to measure running orientation. In the first study (Study 1.1), the baseline accuracy of the popular 6DOF Madgwick filter was assessed for running orientation over the duration of a sports training session (>30 minutes). A linear de-trending function was applied over sections of the Z-axis orientation data, in order to remove the signal drift. This resulted in de-drifted data centered around zero, to which an offset was applied to align this data with the reference. In a laboratory free environment this offset could be calculated before exercise using the average Z-axis Euler orientation angle collected in a few seconds when the person is standing still with the sensor attached to their leg. Overall the results in Study 1.1 remained too high for clinical use, with RMSE greater than 10 degrees on most segment axes. In comparison to previous studies (with errors <6°) [8, 211], it can be seen that this level of error is too high for accurate analysis of running technique. This lead the research to investigate the cause of the high errors and to examine potential error reduction strategies.

The inconsistent errors from Study 1.1 were investigated and it was found that these errors, seemingly random, were not due to signal drift and 6DOF estimation alone, but additionally related to data loss. It was shown that the data loss was greatly affecting the ability of the 6DOF orientation algorithm to converge to the true orientation. To tackle this, the remaining studies in this chapter explored approaches to improve the convergence of the orientation algorithm, which would lead to a reduction in error. These approaches involved (i) investigating the use of a pre-processing pipeline to pre-fill and post-fill the missing data (Study 1.2) and (ii) additionally the exploration of two Madgwick filter modifications (Study 1.3A and Study 1.3B).

Study 1.2 investigated the data loss by attempting to design a pre-processing pipeline for the data. This pipeline involved pre-filling some of the lost samples in the inertial data up to a set threshold (gap-size), and post-filling any remaining lost samples in the orientation data. This was examined with two interpolation approaches: linear interpolation and previous sample interpolation. On average, neither approach was found to be effective in resulting in lower orientation error compared to ignoring the missing data (as in Study 1.1). The implication of this result is that missing inertial data will still pose problems for 6DOF estimation. Future research should examine more advanced gap-filling methods to address this.

In studies 1.3A and 1.3B, modifications to the baseline 6DOF Madgwick filter were examined. The first modification (explored in study 1.3A), optimisation of the beta parameter, has been previously shown in a number of other studies to improve orientation accuracy for both 6DOF [7, 216] and 9DOF [12, 366]. A grid search was performed on a subset of the data to examine a range of values for the beta parameter. This found that beta= 0.1 resulted in the lowest orientation error and was subsequently applied to the full dataset. Although the optimisation of the beta parameter successfully reduced errors compared to the baseline results (Study 1.1), the level of orientation errors were still too high to be used for absolute orientation estimation in a clinical setting.

The second modification, changes to the algorithm design itself was explored in Study 1.3B. These changes were the introduction of multiple iterations of Stochastic Gradient Descent and the inclusion of a momentum parameter. A grid search approach was used to find the combination of parameters the resulted in the lowest orientation error, using beta baseline (B=0.033) and beta optimised (B=0.1). It was found that the best set of parameters was different for each segment, and for each beta parameter tested. Generally, the best parameter combinations resulted in slightly reduced orientation error, however this reduction in orientation error was in generally not significant and had mixed results for the optimised beta set BestParam0.1. Therefore, of the two modification for the orientation algorithm investigated in Study 1.3A and 1.3B, the simple beta parameter optimisation is the recommended approach for future parameter tuning.

Chapter 8

Study 2: Developing a Mobile Solution to Provide Real-time Biofeedback to Reduce Impact Loading during Running

8.1 Introduction

Running is a very popular and accessible form of physical activity [410] that is associated with cardiovascular, musculoskeletal and mental health benefits [411]. However, the incidence rate of running related injuries (RRIs) is high, with between 19 and 80% of runners sustaining a RRI within a 1-year period [5]. As well as the negative health implications this can have on the runner's quality of life (e.g., reduced mobility, pain), sustaining an injury can discourage runners from returning to running [66, 132]. A key factor implicated in RRIs is the high repetitive impact forces which occur when the foot strikes the ground [1, 27], causing high impact forces to travel up through the musculoskeletal system [123, 412].

While previous injury-based research has heavily focused on vertical ground reaction forces (vGRFs) and assessing whole-body loading during running (using a force plate) [1, 224], understanding loading at individual body segments may be important. vGRFs are calculated as the sum of forces occurring across individual body segments, and are not segmentally proportional, meaning there is no information about how the forces are distributed across body segments. Assessing localised loading at individual body segments may provide more relevant information with respect to the relationship between loading and injury, as injuries occur locally at a segment level, and not at a full body level [27]. Small, wearable accelerometer sensors have been more recently used to examine the relationship between segmental loading and both running technique and running injuries, with the vast majority assessing tibial impact accelerations [27, 36].

In light of the relationship between loading and injury, a number of studies have explored the use of real-time biofeedback as a running technique retraining tool to reduce loading and RRIs [23,

224, 225]. Chan et al. [20] used real-time visual feedback from an instrumented treadmill (with embedded force plate) to successfully teach runners to reduce their loading, and in doing so reduced their rate of injury by 62% over a 1-year period. To date, studies have predominantly used real-time biofeedback for running technique retraining based on either measures of loading (e.g. force plate [20] or accelerometers [23]) or joint kinematics (technique) [104]. The application of findings from running technique retraining studies like Chan et al. [20] are somewhat limited due to the majority of retraining studies that provide biofeedback through the use of expensive equipment in constrained laboratory settings with limited access [20, 23, 229]. To further highlight this is issue, van Gelder et al. [25] found that only 4% of their reviewed papers (8/173) provided biofeedback in a non-laboratory environment.

In addition, only one study by Baggaley et al. appears to have directly compared whether loading or technique-based feedback is more effective at providing technique retraining to reduce loading variables during running [24]. This laboratory-based study compared visual feedback based on modified foot strike patterns (technique), modified step length (technique), and vertical average loading rate [VALR] (loading measure via instrumented treadmill with embedded force plate). Baggaley et al. found that all sources of biofeedback resulted in reduction of VALR loading after running retraining, however technique based biofeedback by modifying foot strike patterns (rear-foot strike to non-rear-foot strike) resulted in significantly greater reductions. No studies have directly compared whether biofeedback based on either loading (as measure by impact acceleration) or joint kinematics (technique) is most effective in reducing segmental loading during running. This has neither been examined in a laboratory nor overground running scenario. Such research would provide technology companies and clinicians with valuable information in developing real-time biofeedback technologies and implementing effective running retraining programmes.

Given that the angle of knee flexion at initial contact with the ground is strongly associated with loading on the body, with greater flexion associated with lower loading [26, 116, 413], it would seem appropriate to use knee flexion as the biofeedback measure. However, the analysis of knee flexion angle requires the use of at least two wearable sensors (accelerometer/inertial) which makes the provision of biofeedback more cumbersome and leaves room for alignment errors. Single sensor set-ups may be utilised by calculating segment angles only (1 sensor) as opposed to joint angles. In an attempt to limit the number of sensors involved, preliminary work we have undertaken (not included in this thesis) shows that the thigh angle is a more consistent determinant of knee angle than the tibia angle,

there is less variation in thigh angle at initial contact in comparison to tibia angle. In addition, it would appear that when a person runs 'softer' the thigh angle changed more than the tibia angle. Therefore, this section (Study 1.2) specifically uses thigh angle as the technique-based biofeedback measure. It is unclear whether it is more effective to provide direct biofeedback on loading itself (impact accelerations) or indirect biofeedback (thigh angle). This appears to only be examined in one study, Baggaley et al. [24], with respect to GRF loading (VALR) and no studies using impact acceleration measures of loading. The study by Baggaley et al. suggested that indirect biofeedback on foot strike angle (47% reduction in loading) was more effective that direct biofeedback on VALR (31% reduction in loading). However it is important to recognise that not all indirect measures of biofeedback will be equally effective, as in the same study by Baggaley et al., direct VALR biofeedback (31% reduction in loading) was more effective than indirect step-rate biofeedback (15% reduction in loading). Therefore, more research is required to examine these biofeedback comparisons.

It could be suggested that direct feedback on the variable of interest (impact accelerations) might be more effective than biofeedback on a secondary variable (thigh angle), but no studies have directly compared these specific measures for running retraining. This relates to the question of whether knowledge of performance (thigh angle) or knowledge of results-based biofeedback (impact accelerations) is more effective.

Finally, studies that have focused on the use of biofeedback on segmental loading have predominantly used peak accelerations at the tibia as an outcome measure and ignored measuring loading at other segments. The choice of tibia is most likely because lower limb injuries are common in running [5], the tibia itself is a common site of injury [27], and peak accelerations are an easy measure to extract. However, lower back injuries can also occur [5, 243], suggesting the need to also assess sacrum loading via impact accelerations. In addition, while measures of peak loading are important in understanding RRIs [62], the rate of loading has been shown to be more related to injury at a whole-body level [1]; thereby suggesting the potential importance of understanding the effect of running retraining on segmental rates of acceleration.

The aim of this chapter is to compare two methods of real-time biofeedback, (a) tibia impact accelerations, and (b) thigh angle at initial contact, to reduce peak (Peak_{acc}) and rate (Rate_{acc}) of acceleration measured at the tibia and sacrum. This study will use a smart phone application developed by the researcher as part of this PhD to provide the relevant biofeedback to the runner in real-time.

8.2 Overview of the Data Collection

Section 6.3 describes the datasets (two groups) which were used in this Chapter 8, and how the data were collected. The data collection protocol for each dataset was identical, differing only in the type of biofeedback that was provided to the runner, and where the biofeedback sensor was places on the body. Biofeedback based on impact accelerations were provided to Group A via the tibia, and biofeedback based on thigh angle was provided to Group B via the thigh. Sixteen volunteers were recruited for each of the two groups. The experiment involved participants running in an outdoor environment, while wearing three wireless Shimmer3 inertial sensors (Shimmer Sensing, Dublin, Ireland) attached to their body. Two inertial sensors were placed on the sacrum and tibia, these were to assess loading on the body. The third sensor was placed on either the tibia or the thigh, to provide biofeedback based on impact accelerations or thigh angle, respectively. Full experimental set-up is described in section 6.3.

8.3 App Development and Design

Auditory biofeedback was provided to a runner via a mobile phone application. No similar mobile application appeared to exist already publicly; therefore, mobile phone application (app) development was required as part of this work. A requirement for this app was the ability to handle sensor data in real-time. The app was developed by the author of this thesis through the use of the Shimmer Sensing API and using the foundations of a simple Bluetooth connectivity app made available by Shimmer Sensing. This Bluetooth connectivity app had the base functionality that was required to connect a wireless Shimmer3 inertial sensor (Shimmer Sensing, Dublin, Ireland) to a mobile phone app for real-time streaming of inertial data via Bluetooth. On top of this foundation, multiple functions were added to the app to fit the requirements of the present study.

To set-up the app, the biofeedback sensor was switched on and paired to the phone via Bluetooth. From the app menu, the user could select which sensors (accelerometer, gyroscope) to enable and stream in real-time. There were two versions of the app designed to provide different sources of biofeedback. Version 1, where the accelerometer was enabled, provided biofeedback based on impact accelerations and Version 2 where both the accelerometer and gyroscope were enabled, provided biofeedback based on thigh angle. In its current implementation, switching between app versions (and the real-time biofeedback provided) was as simple as disabling/re-enabling the

gyroscope sensor in-app. Once streaming, the user could toggle a recording button which would save all enabled inertial data to the phone, for offline analysis. Depending on which sensors were enabled, accelerometer impact acceleration data (accelerometer enabled) or thigh angle data (gyroscope and accelerometer enabled) were plotted in real-time on the display within the app for visualisation purposes.

The app determined when the user started running and after a few seconds of running, the user was able to record baseline values of the measure that was to be reduced (either impact accelerations or thigh angle). When the user was finished recording their baseline run, the maximum peak impact acceleration (Group A) or minimum thigh angle at initial contact (Group B) during that period was prompted on the screen, this was the value that was representative of the baseline run. With only the accelerometer enabled, peak tibial accelerations were extracted from the raw acceleration data streaming in real-time to the app. With the accelerometer and gyroscope enabled, 6-Degrees of Freedom (6DOF) thigh segment angle data was calculated within the app using the beta optimised Madgwick filter (from Section 7.5) in real-time.

This provision of feedback via the mobile phone was designed to reduce baseline peak impact accelerations by 10% and thigh angle at initial contact by 5 degrees. The app allows setting of this threshold, which was visually displayed on the graph along with the streaming data. Once a threshold was set, the app gave simple audio tone 'beep' feedback in response to exceeding the set threshold. Feedback was not instantaneous but was provided within a second of exceeding the threshold (near real-time).

8.4 The Provision and Comparison of Loading vs Technique based Biofeedback during Running

With respect to loading, previous running retraining research has proven the effectiveness of loading based biofeedback at providing reductions in loading on the body, measured as impact accelerations via accelerometers [23, 226, 229] or as vGRF measures (VALR, VILR, VIP, [20, 24, 228]) via instrumented treadmills (with force plates) [see tables 4.2 and 4.1]. Running retraining studies appear to be capable at providing a minimum reduction of approximately 10% across loading variables [see table 4.2]. The majority of studies examining loading based biofeedback, provide biofeedback via tibial impact accelerations but the effectiveness of running retraining programmes at reducing loading at non-tibial

segments does not appear to be well researched [26]. It is important to understand how the changes in technique may affect loading further up the body e.g., at the sacrum, where injuries may occur [243].

Technique-based biofeedback has also been shown to be effective at reducing loading on the body [24, 224, 225], however only one study could be found which compared the relative effectiveness of technique based biofeedback and loading based biofeedback in Baggaley et al. [24]. Baggaley et al. investigated the reduction of loading associated with the provision of loading based biofeedback (VALR) and technique based biofeedback (foot strike pattern, step rate) during treadmill running. They found both to sources of biofeedback be effective, but technique based biofeedback more so. Due to the limited research, it is unclear which type of biofeedback is more effective at reducing loading on the body during running.

Research has suggested that the role of knee kinematics at initial contact plays an important role in the attenuation of impact forces on the body [414]. Specifically increased knee flexion on landing can greatly reduce the magnitude of vertical impact forces [122]. In this thesis, technique based biofeedback was investigated through providing runners with biofeedback related to thigh angle. The thigh angle may be useful for biofeedback as pilot work carried out by the researcher has shown that the angle of the thigh remains more consistent at initial contact in comparison to the tibia, and may be more important at influencing reductions in loading when running softer.

8.4.1 Methodology

8.4.1.1 Data Processing

Section 6.3 includes a detailed description of the data collection and study design. A mobile phone application (app) was designed to allow for running retraining with auditory biofeedback in an outdoor environment. In order to provide the biofeedback, the mobile app was required to connect to a Shimmer3 inertial sensor (Shimmer Sensing, Dublin, Ireland) and stream inertial data in real-time via Bluetooth. With the app implemented and tested, volunteers were recruited for a running retraining study.

The study was designed with three main phases which are described in detail in 6.3. During Phase One of normal running (no biofeedback provided), data were recorded on the loading assessment sensors of the sacrum and tibia, measuring local segment acceleration via accelerometers. During this time, baseline values for the biofeedback variable, either (i) peak impact acceleration or (ii) thigh angle at initial contact, were recorded on the mobile phone app. This baseline value was used to set

a fixed threshold with-in the app to active biofeedback during Phase 2 and 3. In Phase Two, no data were collected on the app or the loading assessment sensors, this phase allowed the runner to become familiar with the auditory biofeedback provided by the app. During Phase Three, data were recorded on the loading assessment sensors while the runner aimed to 'run softer'. The goal for the runner was to reduce their baseline value of peak impact acceleration or thigh angle at initial contact (as recorded in Phase 1) by 10% or 5 degrees, respectively. If the measured biofeedback variable of the runner exceeded the set threshold, the audio biofeedback sounded a 'beep' tone from the phone to alert the runner to 'run softer', and that they were not staying below their set threshold.

For each segment (tibia and sacrum) ten strides of running were extracted from both loading data measured during running (without biofeedback [Phase One] and with biofeedback [Phase Three]) at the 5-minute mark. Average peak impact acceleration and average rate of acceleration across the ten strides for each running type were calculated for each segment. This analysis was to evaluate whether biofeedback reduced loading on the tibia and sacrum by a significant reduction in either the peak impact acceleration or rate of acceleration. Peak impact accelerations and the rates of acceleration were calculated using the Y-axis accelerometer sensor data, as this was aligned vertically with the body (and gravity) on each segment, and therefore had the largest magnitudes of acceleration.

8.4.1.2 Data Analysis

Statistical analysis was carried out for each dependent variable, peak impact acceleration ($Peak_{acc}$) and rate of acceleration ($Rate_{acc}$), using paired t-tests between the two groups of running types: normal and biofeedback running. This analysis was carried out separately for the tibia and sacrum data. In order to compare the effectiveness of the two types of biofeedback, thigh angle and impact accelerations, and the effect of gender, three-way mixed ANOVAs were carried out for each dependent variable ($Peak_{acc}$ and $Rate_{acc}$). For each of these analyses 'Running Type' (Pre-Feedback vs With-Feedback) was a within groups factor, and 'Biofeedback Type' (impact accelerations or thigh angle) and 'Gender' were between groups factors. The level of significance was set at $p < 0.05$. The effect size was measured using (partial) eta squared (η^2), where $\eta^2 = 0.01$ indicated a small effect, $\eta^2 = 0.06$ indicated a medium effect and $\eta^2 = 0.14$ indicated a large effect size. The statistical analysis was completed using IBM SPSS Statistics 24.

8.4.2 Results

There were no outliers in any of the acceleration data, as assessed by inspection of boxplots for values greater than 1.5 box-lengths from the edge of the box. All acceleration measures were normally distributed, as assessed by Shapiro-Wilk's test ($p > .05$). There was homogeneity of variances for all acceleration measures, as assessed by Levene's test for equality of variances. Summary data (mean $\pm$ standard deviation) for each condition are provided in Tables 8.1 and 8.2. Percentage changes due to FeedbackType are also provided; these were calculated by combining Genders.

Table 8.1: Summary data for impact accelerations for the tibia acceleration based feedback group.

	Tibia-Acceleration based Feedback Group						
	Pre-Feedback			With-Feedback			Percentage change
	Male	Female	Gender Combined	Male	Female	Gender Combined	Pre- to With-Feedback[1]
	$(\overline{x}\pm$ SD)	$(\overline{x}\pm$ SD)	$(\overline{x}\pm$ SD)	$(\overline{x}\pm$ SD)	$(\overline{x}\pm$ SD)	$(\overline{x}\pm$ SD)	$(\overline{x}, 95\%CI)\%$
Tibia Peak$_{acc}$ (g)	-18.96 (±3.04)	-20.19 (±2.33)	-19.57 (±2.69)	-15.26 (±2.62)	-16.11 (±2.57)	-15.68 (±2.55)	19.9% (17.3, 22.7)
Sacrum Peak$_{acc}$ (g)	-4.35 (±0.63)	-3.74 (±0.93)	-4.04 (±0.83)	-3.77 (±0.58)	-3.25 (±0.80)	-3.51 (±0.73)	13.2% (12.3, 14.2)
Tibia Rate$_{acc}$ (g/s)	2551.6 (±775.0)	2632.5 (±460.6)	2592.1 (±617.3)	2166.8 (±670.4)	2229.6 (±441.6)	2198.2 (±549.4)	15.3% (13.5, 17.1)
Sacrum Rate$_{acc}$ (g/s)	219.6 (±72.2)	184.1 (±54.9)	201.9 (±64.6)	197.2 (±65.8)	166.1 (±49.6)	181.6 (±58.6)	10.1% (9.6, 10.6)

Note: [1] percentage change Pre-feedback to With-feedback is for Gender-combined as there was no statistically significant difference due to gender.

Table 8.2: Summary data for impact accelerations of the thigh angle based feedback group.

	Thigh Angle based Feedback Group						
	Pre-Feedback			With-Feedback			Percentage change
	Male	Female	Gender Combined	Male	Female	Gender Combined	Pre- to With-Feedback[1]
	$(\overline{x}\pm$ SD)	$(\overline{x}\pm$ SD)	$(\overline{x}\pm$ SD)	$(\overline{x}\pm$ SD)	$(\overline{x}\pm$ SD)	$(\overline{x}\pm$ SD)	$(\overline{x}, 95\%CI)\%$
Tibia Peak$_{acc}$ (g)	-18.51 (±2.11)	-19.71 (±0.77)	-19.12 (±1.65)	-16.55 (±2.01)	-18.43 (±1.60)	-17.49 (±2.00)	8.6% (5.3, 11.8)
Sacrum Peak$_{acc}$ (g)	-3.94 (±0.71)	-4.22 (±0.64)	-4.08 (±0.67)	-3.15 (±0.54)	-3.29 (±0.58)	-3.22 (±0.55)	20.5% (15.2, 25.7)
Tibia Rate$_{acc}$ (g/s)	2640.6 (±636.9)	2523.9 (±503.7)	2582.3 (±558.0)	2391.0 (±552.2)	2319.7 (±383.1)	2355.3 (±460.6)	8.2% (4.3, 12.2)
Sacrum Rate$_{acc}$ (g/s)	234.7 (±45.9)	208.6 (±50.4)	221.6 (±48.5)	194.9 (±46.4)	176.9 (±57.0)	185.9 (±61.7)	16.1% (9.6, 22.6)

Note: [1] percentage change Pre-feedback to With-feedback is for Gender-combined as there was no statistically significant difference due to gender.
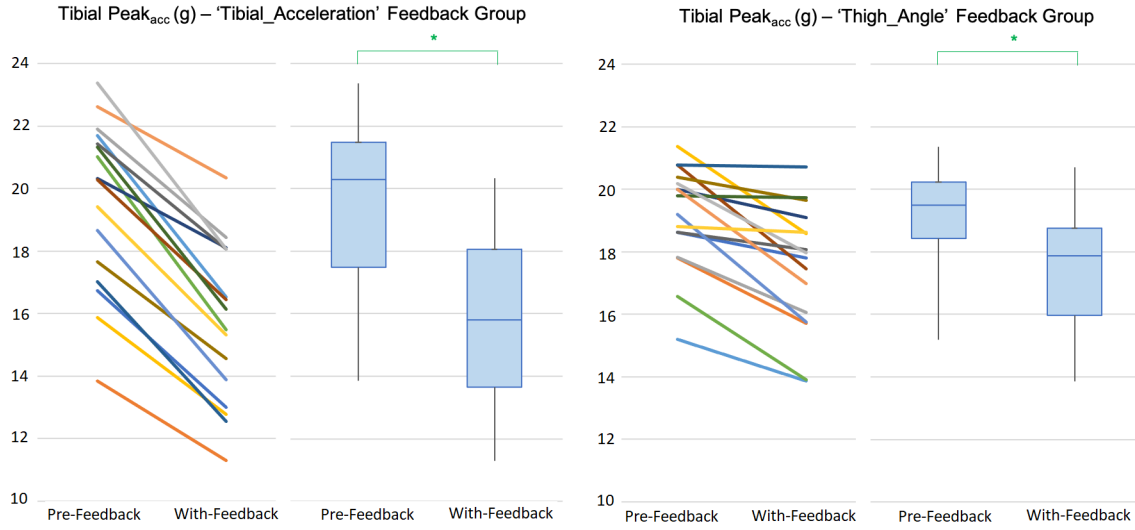
Figure 8.1: The effect of FeedbackType (ThighAngle vs TibiaAcc) on Peak Impact accelerations: individual participants (left) and group box-plots for Pre-Feedback and With-Feedback. Note: Peak$_{accel}$ values are displayed as positive values for visualisation purposes. Males and females are combined as there were no significant differences. * Significantly reduced due to feedback.

The findings for all four acceleration measures (Tibia Peak$_{accel}$, Tibia Rate$_{accel}$, Sacrum Peak$_{accel}$, Sacrum Rate$_{accel}$) followed similar overall patterns of significance; no significant three-way interactions were found, significant two-way interactions between Feedback (Pre-feedback vs With-feedback) and FeedbackType (TibiaAcc-based feedback, ThighAngle-based feedback) were evident, with both feedback types producing significant reductions in accelerations (see figures 8.6 and 8.5). Effects of FeedbackType at an individual participant level are given in figures 8.1 and 8.2 for Peak$_{accel}$, and figures 8.3 and 8.4 for Rate$_{accel}$. TibiaAcc-based feedback appears to result in more consistent reductions in accelerations across participants than ThighAngle-based feedback.

Tibia Peak Acceleration (Tibia Peak$_{accel}$)

There was no statistically significant three-way interaction between Feedback, Gender and FeedbackType, $F(1, 28) = 1.685$, p= 0.205, partial η^2= .057. There was a statistically significant two-way interaction between Feedback and FeedbackType, $F(1, 28) = 31.53$, p < .0005, partial η^2 = .530. No main effects of Gender or other with-in or between two-way interactions were statistically significant (p>0.05).

Post-hoc paired t-test analyses of the effect of Feedback, indicated that both TibiaAcc-based feedback and ThighAngle-based feedback resulted in a significant reduction in tibia Peak$_{accel}$ (t = -14.14, p< 0.001, Cohen's d = -1.48, diff% = 20.0%; t = -5.45, p<0.001, Cohen's d = -1.70, diff%

180
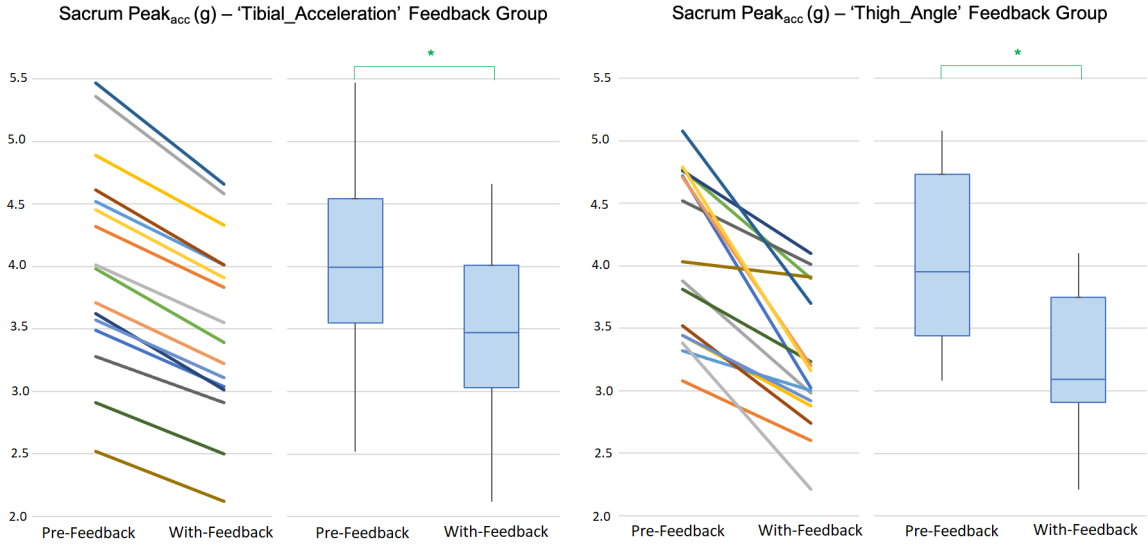
Figure 8.2: The effect of FeedbackType (ThighAngle vs TibiaAcc) on Peak Impact accelerations: individual participants (left) and group box-plots for Pre-Feedback and With-Feedback. Note: Peak$_{accel}$ values are displayed as positive values for visualisation purposes. Males and females are combined as there were no significant differences. * Significantly reduced due to feedback.

= 8.5%, respectively). Furthermore, an independent t-test found the reductions were statistically significantly larger for the TibiaAcc-based feedback than the ThighAngle-based feedback (t = -5.59, p< 0.001, Cohen's d = -1.98).

Sacrum Peak Acceleration (Sacrum Peak$_{accel}$)

There was no statistically significant three-way interaction between Feedback, Gender and FeedbackType, $F(1, 28) = 0.777$, p = .386, partial η^2 = .027. There was a statistically significant two-way interaction between Feedback and FeedbackType, $F(1, 28) = 6.38$, p = .017, partial η^2 = .186. No main effects of Gender or other with-in or between two-way interactions were statistically significant (p > .05).

Post-hoc paired t-test analyses of the effect of Feedback, indicated that both TibiaAcc-based feedback and ThighAngle-based feedback resulted in a significant reduction in sacrum Peak$_{accel}$ (t = -17.09, p < 0.001, Cohen's d = -0.68, diff% = 13.2%; t = -7.06, p < .001, Cohen's d = -0.75, diff% = 20.5%, respectively). Furthermore, an independent t-test found the reductions were statistically significantly larger for the ThighAngle-based feedback than the TibiaAcc-based feedback (t = -2.57, p = 0.015, Cohen's d = 1.06).
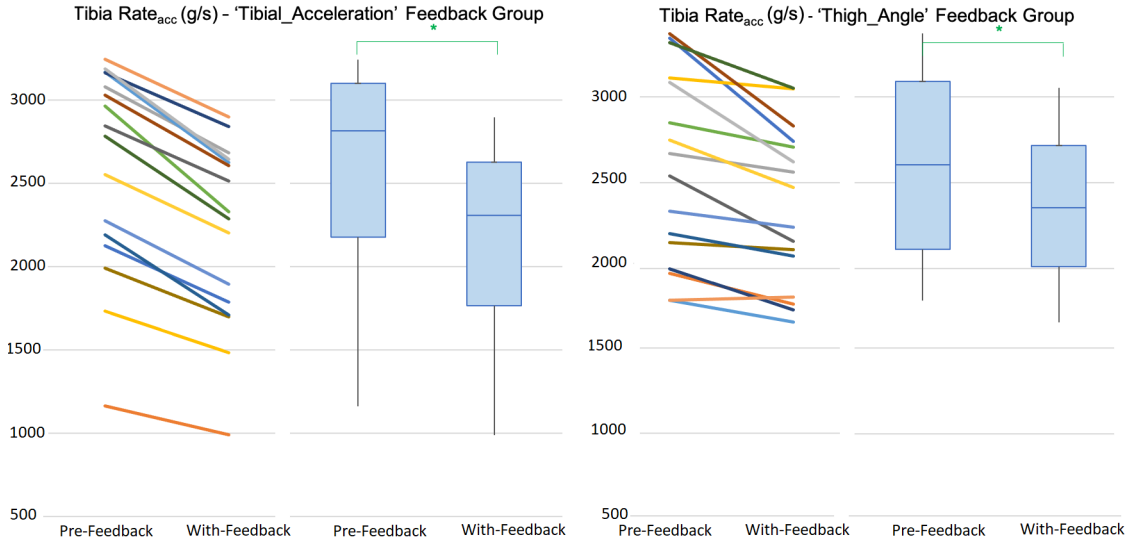
Figure 8.3: The effect of FeedbackType (ThighAngle vs TibiaAcc) on Rate of acceleration: individual participants (left) and group box-plots for Pre-Feedback and With-Feedback. Males and females are combined as there were no significant differences. * Significantly reduced due to feedback.

Tibia Rate Acceleration (Tibia Rate$_{accel}$)

There was no statistically significant three-way interaction between, Gender and FeedbackType, $F(1, 28) = 0.317$, $p = 0.578$, partial $\eta^2 = 0.011$. There was a statistically significant two-way interaction between Feedback and FeedbackType, $F(1, 28) = 8.70$, $p = 0.006$, partial $\eta^2 = 0.239$. No main effects of Gender or other with-in or between two-way interactions were statistically significant ($p > .05$).

Post-hoc paired t-test analyses of the effect of Feedback, indicated that both TibiaAcc-based feedback and ThighAngle-based feedback resulted in a significant reduction in tibia Rate$_{accel}$ ($t = 13.11$, $p < 0.001$, Cohen's d = 0.67, diff% = 15.3%; $t = 4.95$, $p < 0.001$, Cohen's d = 0.83, diff% = 8.3%, respectively). Furthermore, an independent t-test found the reductions were statistically significantly larger for the TibiaAcc-based feedback than the ThighAngle-based feedback ($t = 3.05$ $p = 0.005$, Cohen's d = 1.10).

Sacrum Rate Acceleration (Sacrum Rate$_{accel}$)

There was no statistically significant three-way interaction between Feedback, Gender and FeedbackType, $F(1, 28) = 0.073$, $p = .788$, partial $\eta^2 = .003$. There was a statistically significant two-way interaction between Feedback and FeedbackType, $F(1, 28) = 4.9$, $p = .035$, partial $\eta^2 = .149$. No main effects of Gender or other with-in or between two-way interactions were statistically significant ($p > .05$).
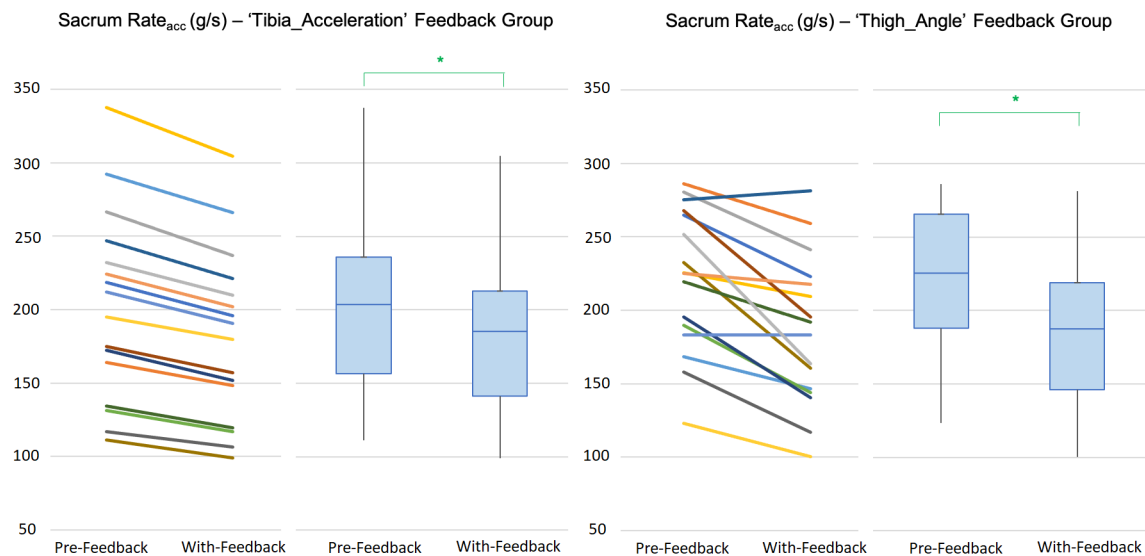
Figure 8.4: The effect of FeedbackType (ThighAngle vs TibiaAcc) on Rate of acceleration: individual participants (left) and group box-plots for Pre-Feedback and With-Feedback. Males and females are combined as there were no significant differences. * Significantly reduced due to feedback.

Post-hoc paired t-test analyses of the effect of Feedback, indicated that both TibiaAcc-based feedback and ThighAngle-based feedback resulted in a significant reduction in sacrum $\text{Rate}_{\text{accel}}$ (t = 12.74, p < 0.001, Cohen's d = 0.32, diff% = 10.1%; t = 5.36, p < .001, Cohen's d = 0.34, diff% = 16.1%, respectively). Furthermore, an independent t-test found the reductions were statistically significantly larger for the ThighAngle-based feedback than the TibiaAcc-based feedback (t = -2.56, p = 0.032, Cohen's d = -0.94).
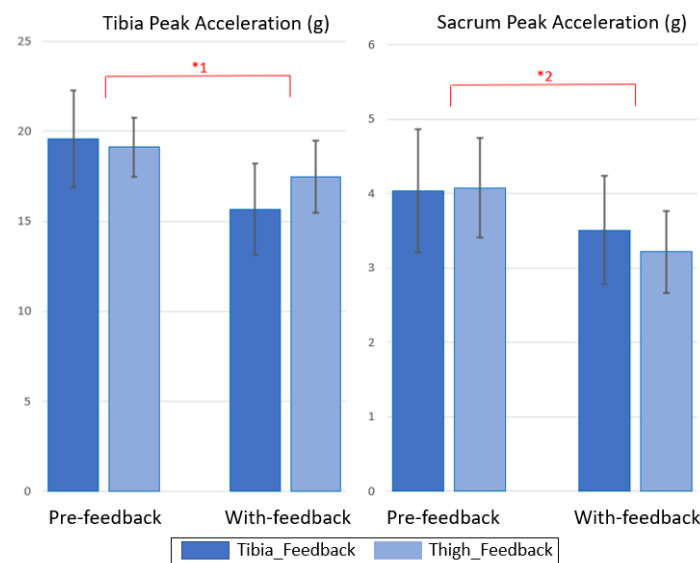
Figure 8.5: The effect of Feedback (Pre-Feedback vs With-Feedback) and FeedbackType (ThighAngle vs TibiaAcc) on peak acceleration (means ± standard errors). Note: Peak values are displayed as positive values for visual presentation. Males and females are combined as there were no significant differences. * Significantly reductions due to Feedback, *1 Reductions due to Feedback is significantly greater with TibiaAcc than ThighAngle, *2 Reductions due to Feedback is significantly greater with ThighAngle than TibiaAcc
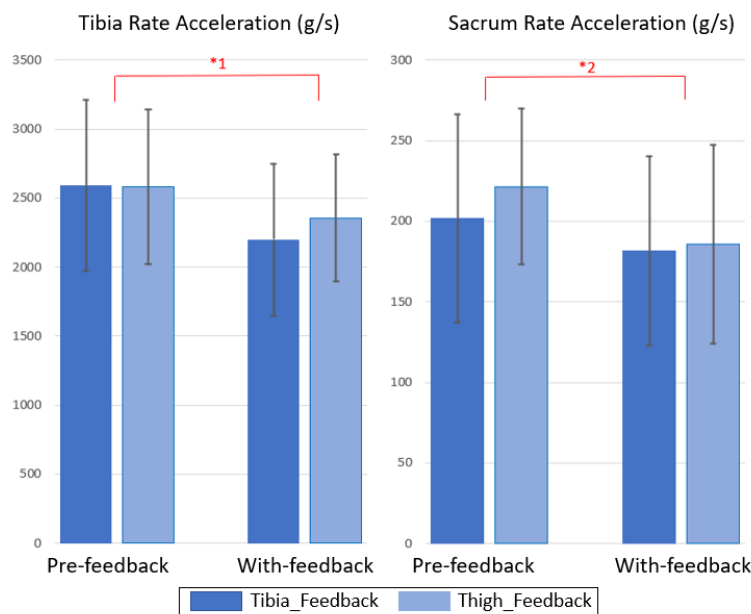
Figure 8.6: The effect of Feedback (Pre-Feedback vs With-Feedback) and FeedbackType (ThighAngle vs TibiaAcc) on rate of acceleration (means ± standard errors). Males and females are combined as there were no significant differences. * Significantly reductions due to Feedback, *1 Reductions due to Feedback is significantly greater with TibiaAcc than ThighAngle, *2 Reductions due to Feedback is significantly greater with ThighAngle than TibiaAcc

8.4.3 Discussion

The aim of this study was to examine and compare the effectiveness of two different types of biofeedback [impact acceleration and thigh angle based biofeedback] at reducing loading on the body during running, measured via acceleration variables at the tibia and sacrum. The results are in support of the hypotheses that both impact acceleration and thigh angle-based biofeedback lead to statistically significant reductions in the loading variables ($Peak_{accel}$ and $Rate_{accel}$) at the tibia and sacrum. Several have previously used either impact acceleration based biofeedback [23, 226, 229] or other forms of technique based biofeedback (step rate [248], foot strike pattern [24], hip kinematics [225]) to reduce measures of loading (vGRF, impact accelerations) on the body.

With a focus on studies where local segmental loading was measured via accelerometers [Tibia $Peak_{accel}$, Sacrum $Peak_{accel}$], and **impact accelerations** were provided as biofeedback; Van den Berghe et al. [235] and Van den Berghe et al. [234] found significant reductions in peak tibial impact accelerations of approximately 26%, when using auditory tibial impact peak feedback during overground running. These results were better than the significant reductions in peak impact acceleration found in this thesis of 20% and 13% for the tibia and sacrum, respectively. Cheung et al. [242] reported slightly greater reduction in peak tibial impact acceleration using auditory biofeedback during treadmill running. They found 21% (non-significant) and 34% (significant) reductions in tibia and foot impact peaks, respectively with the use of biofeedback. The difference in reduction magnitudes across studies are likely related to the initial objective for each runner in each respective study, in Cheung et al. [242] runners were asked to aim to reduce their tibial impact peaks by 20%, in Van den Berghe et al. [235] and Van den Berghe et al. In the work by Van den Berghe et al. [234] runners aimed to reduce their peak tibial impact acceleration by 30% and 50%, respectively, and runners in all studies were encouraged to keep to the threshold using biofeedback. These thresholds are much larger than the 10% threshold on peak impact acceleration used in this thesis, which explains why their runners enacted larger reduction magnitudes. Additionally, each of these studies (Van den Berghe et al. [234], Van den Berghe et al. [235] and Cheung et al. [242]) screened participants in advance of the running retraining and only participants with high peak tibial acceleration were specifically included. It is likely that this may also have led to inflated magnitudes of load reduction with biofeedback as all participants had already higher-than-expected peak tibial acceleration. No such screening was carried out in this thesis. In a separate study, Wood and Kipp [239] provided auditory peak tibial acceleration based feedback during treadmill running

and found a significant reduction of 10-12% in loading (measured as peak tibial impact acceleration) when aiming to reduce tibial impact peaks by a threshold of 10-15%. Wood and Kipp [239] is most related to the research in this thesis in the threshold used (10% vs 10-15%) and the lack of participant screening (mix of high and low loading participants). The significant reductions in peak impact acceleration found in this thesis were greater (20% and 13% for the tibia and sacrum, respectively) then that found in Wood and Kipp [239].

With regard to studies where local segmental loading was measured via accelerometers [Tibia Peak$_{accel}$, Sacrum Peak$_{accel}$], and **technique** was provided as biofeedback; It is difficult to compare findings as it appears no studies have directly examined the effect of thigh angle based biofeedback. In addition, by observation it seems most running retraining studies which use technique based biofeedback, mainly focus on other outcome parameters such as vGRF loading measures (VIP, VALR, VALR) or measures of changes in joint kinematics. Huang et al. [244] examined the effect of visual and auditory technique based biofeedback (combination of foot strike patterns and step rate modification) reported a reduction in peak tibial impact acceleration of 8-13% which is in line with the results in this thesis, which found significant reductions in peak impact acceleration (due to thigh angle biofeedback) of 9% and 21% for the tibia and sacrum, respectively.

No studies appear to have measured the effect of impact loading or technique based biofeedback on the rate of loading at a segmental level, as assessed through the rate of acceleration. This thesis showed a significant level of reduction of 15% and 10% for the Tibia$_{rate}$ and Sacrum$_{rate}$, respectively, using impact acceleration biofeedback, and a significant reduction of 8% and 16% for the Tibia$_{rate}$ and Sacrum$_{rate}$, respectively, using thigh angle biofeedback. The significant reduction in loading reported in this thesis may in fact be sufficiently reduced in order to prevent a RRI. Although this is difficult to say as this was an acute form of running retraining where biofeedback was not maintained or injuries tracked post study. However, Chan et al. [20] showed a reduction of approximately 18% in vGRFs loading rate (VALR and VILR) in the running-retraining group, resulted in a 62% lower injury risk compared with controls. In order to compare to vGRF rate, our reductions in average rate of acceleration (average of tibia and sacrum) were (Table 8.1 and 8.2): 12.7% (TibiaAcc feedback) and 12.2% (ThighAngle feedback). Our average rate of acceleration across participants is lower than the 18% achieved in Chan [20], however, it can be seen from table 8.1 that some individual participants were able to achieve lower reduction in line with Chan [20]. Additionally, comparisons can be made to other studies evaluating running injuries versus healthy controls; Milner et al. [27] examined the

differences in vGRFs (peak and rate) between tibial stress fracture patients and healthy controls. It can be seen from their measured levels of loading between the two groups that the average vGRF rate (VALR and VILR) is 15% lower in healthy controls compared to the baseline of patient with tibial stress fracture. Some studies (Hunter et al. [245] and Cheung et al. [224]) highlighted technique retention due to auditory based biofeedback, where new running techniques were maintained from a month [245] to 3-months [224] post-training. It is possible that the magnitude of reduction in this thesis may support the prevention of future running related injuries.

The support of the sub-hypothesis that loading based biofeedback would be more effective at reducing loading on the tibia and sacrum, is not as clear. While both feedback types were effective in reducing $Peak_{accel}$ and $Rate_{accel}$ at both the tibia and sacrum, the type of biofeedback provided influenced the magnitude of effect. Tibia acceleration-based feedback resulted in significantly greater reductions in $Peak_{accel}$ and $Rate_{accel}$ at the tibia, while thigh angle biofeedback resulted in significantly greater reductions in $Peak_{accel}$ and $Rate_{accel}$ at the sacrum (see figures 8.5 and 8.6). Given that the type of biofeedback in our study differed in two ways, the location of where the biofeedback sensor was placed (tibia versus thigh) and the source of information (tibia acceleration versus thigh angle), it is not possible to conclude on which of these elements most influenced the greater enhancements. It has been suggested that the type of biofeedback, whether knowledge of performance or knowledge of technique based biofeedback is provided, may be important. However, there is a succinct lack of research directly comparing both forms of biofeedback during running, where Baggaley et al. found knowledge of performance, i.e., technique based biofeedback, to be more effective at reducing VALR [24]. In contrast it appears in this thesis knowledge of results-based biofeedback (impact accelerations) performed better on average across the reductions in peak and rates for the tibia and sacrum (see table 8.1). It is unclear why our knowledge of results-based feedback was more effective on average than knowledge of technique; it may be due to the location-based differences or to the single-session, acute nature of the study.

Reductions in loading on the body can be brought about by different changes in technique at initial contact such as increased flexion at the knee and hip [246], changing from rear-foot strike to non-rear-foot strike and increasing ankle flexion [224], increased step rate [248], lower centre of mass and lower foot vertical velocity [250]. It is possible that the thigh-based feedback resulted in more knee and hip flexion at initial contact, as access to knee kinematics were known while running. Furthermore, while tibia-based feedback also was effective at reducing impact loading but had no access to knee

kinematics, tibia-based feedback may have resulted in more in ankle dorsi-flexion, and less centre of mass and foot vertical velocity at initial contact. However, given that the actions of all joints and the vertical velocity of the centre of mass and the foot effect both biofeedback measures (impact acceleration and thigh angle) at the tibia and thigh, as well as the outcome loading/acceleration measures at the tibia and sacrum, future research should directly examine how the different types and location of feedback effect running technique.

Interestingly, the reduction in acceleration (peak and rate) at both the tibia and sacrum were more consistent across participants when provided by acceleration-based feedback from the tibia than by angle-based feedback from the thigh (see figures 8.1, 8.2, 8.3 and 8.4). It is not possible to conclude if this difference is due to the effect of feedback type (impact acceleration versus thigh angle) or location (tibia versus thigh), but it is possible that knowledge of results (i.e., peak impact acceleration) based feedback is more effective in this regard. Future research should examine this directly. It is worth noting however, that in examining how runners control changes in knee angle, our pilot research found that larger and more consistent changes were produced at the thigh than the tibia. This may make the use of angle-based feedback from the tibia more problematic to use than from the thigh, as was our initial justification in using the thigh for angle-based feedback in the present study. For this reason, we would suggest researchers initially compare acceleration-based feedback measured at the tibia and at the thigh, rather than compare angle-based feedback at these segments.

From a practical applications perspective, it has been shown that both biofeedback based on tibial impact acceleration and thigh angle are effective in reducing both peak and rate of acceleration at the tibia and sacrum. Given that it is easier to secure an inertial sensor to the tibia than the thigh, biofeedback based on tibial impact acceleration may be easier to provide. Although the location of the injury being targeted will primarily influence the decision on which sensor location to use, the greater number of RRIs occurring in the lower limbs compared to the hip and lower back may further support the use of tibia based biofeedback [20].

8.4.4 Future Work

Given that the type of biofeedback in our study differed in two ways, the location where biofeedback was provided (tibia versus thigh) and the source of information (acceleration versus angle), it is not possible to conclude which of these elements most influenced the greater reductions in impact loading on the body. Future research should directly examine how the different types of feedback and location

of feedback may affect running technique and running retraining outcomes. Auditory biofeedback has been shown to promote technique retention [224, 245], where new running techniques were maintained from a month [245] to 3-months [224] post-training. It would be interesting to know what the effect of our portable biofeedback running retraining app would have on technique retention and injury risk reduction, and how often would retraining be required in order to maintain more compliant running strategies.

8.4.5 Conclusion

Both acceleration-based biofeedback from the tibia and technique-based biofeedback from the thigh are effective in reducing peak and rate of acceleration at both the tibia and sacrum. However, TibiaAcc-based feedback resulted in significantly greater reductions in $Peak_{accel}$ and $Rate_{accel}$ at the tibia, while ThighAngle-based feedback resulted in significantly greater reductions in $Peak_{accel}$ and $Rate_{accel}$ at the sacrum. Given the higher number of injuries at the knee and lower leg compared to the hip and lower back [20], as well as the greater ease in securing the tibia-based sensor, we recommend the preferential use of tibia-based feedback over thigh-based feedback with all else being equal. In light of the ability of biofeedback based running re-education programmes to reduce the likelihood of running related injuries [20], future studies should explore the effectiveness of inertial sensors to reduce running related injuries using acceleration- or angle-based biofeedback.

Chapter 9

Study 3: Developing Machine Learning Models for the Prediction of Running-Related Injuries using Impact Loading and Running Technique

9.1   Introduction

Many studies in the area of biomechanics and injury aim to examine the underlying relationships between injury and a number of biomechanical and non-biomechanical variables (see studies outlined in Section 2.2 of this thesis examining potential risk factors of RRI's). When significant relationships are identified between the variables and injury, they provide direction for applied intervention strategies and future intervention studies. While intervention strategies can be applied to all runners, it can be challenging to persuade all runners to undertake an injury prevention programme. Even adherence to simple injury prevention strategies may be poor [36]. For example, a study which recruited 45 runners as part of an examination of muscular tightness found 27% of the runners said that they never perform stretching exercises [415]. A clear alternative approach could be to first identify those runners who are most likely to become injured and target these runners specifically for interventions. Injury prevention interventions have been shown to be helpful in re-training running technique and ultimately reducing the risk of running related injuries [20]. The ability to inform runners that they may have an increased likelihood to incur a running injury due to their movement technique or loading would be extremely helpful. By only focusing on those who are more likely to become injured this is more beneficial to all runners [416], and uptake of injury prevention strategies by runners may be higher due to the potential benefits and motivation for this group of runners in particular.

These are two similar but differing approaches in this regard, one tries to find relationships between informative variables and injury via inferential statistics (to find variables which can be targeted for intervention studies) and the other approach aims to create a model based on any selection of variables to predict likelihood of injury [417]. With prediction, although the variables explored will mainly be

informed by the research domain of running related injuries, the variables themselves are not of concern, the primary objective is the binary outcome of (likelihood of) injury or not.

The work in this thesis focuses on a prediction task to identify those runners who may be predisposed to a general running related injury (RRI). This chapter explores using various supervised machine learning classifiers to attempt to build a reliable injury prediction model. This is a supervised classification problem with the objective of classifying runners as belonging to the injured or non-injured groups. Using a dataset of prospectively injured and non-injured runners, a set of input variables (features) describing the data are assigned a label as either injured or non-injured, and are used to define a mapping function on which a machine learning model is trained. The model is tested with a set of unseen data (from runners whose data are not involved in building the model) and the model then makes a prediction using the unseen runner's data, classifying each runner as mapping to either the injured or non-injured group. Machine learning could be able to infer relationships between the data of the injured and non-injured runners which may not be explicitly clear, thus providing a system that will be able to make future predictions on whether the data from new unseen participants map to injured or non-injured groups. If the machine learning model can classify an unseen participant (runner) to the 'injured' group, this suggests that something about their running data (loading or technique) is associated with the data of those who prospectively became injured.

From the literature review in Section 2.2.3, it is apparent that a number of retrospective studies have examined the relationship between injury and loading, or between injury and technique, but there is a need for more large-scale prospective research. With regards to the use machine learning for the prediction of injury, while using either inertial sensor and/or kinematic technique data, 17 studies were identified in total, where 10/17 studies focused on the prediction of specific injuries (primarily ACL injuries), and the remaining 7 studies investigated general RRIs. The accuracy in terms of AUC-ROC ranged from 0.51-0.8 for these studies. There is a general held belief that more work needs to be done on identifying useful features for machine learning models to predict RRIs.

9.2 Machine Learning: General Running Injury Predictor Model

9.2.1 Aim

The main aim of this study was to compare different datasets using inertial data (accelerometer and gyroscope), movement technique data (joint angles) and clinical data in determining which combination or dataset is more effective at identifying those who are predisposed to RRI. Machine learning could be able to accurately predict a runner's likelihood of injury based on a number of variables related to their impact loading and technique (kinematics) during running and their clinical data. Basic clinical data (age, sex, height, weight, body mass index, previous injury, running experience, training load) were included as these have been shown to be confounding factors and (intrinsic or extrinsic) risk factors for RRIs (see section 2.2.4). Within this question we are particularly interested in whether there is an added benefit in the inclusion of technique-based features with inertial features, versus inertial features only. As we have shown it isn't straightforward to collect accurate technique data in the wild, therefore an accurate predictor system based on inertial features alone would be very beneficial. This type of analysis would allow future injury prevention techniques to be targeted at a subset of runner who would be identified as of at risk of injury. Those runners who are flagged as 'likely to become injured' by the injury prediction model. Previous studies support the relationship between technique and RRI's [29, 94] but more research is needed. Clinical data features were included as these variables represent intrinsic and extrinsic risk factors for RRI's (see Section 2.2.4). A final sub-aim to identify injury causative factors and to determine whether a specific machine learning classifier was more effective at distinguishing those predisposed to injury.

9.2.2 Methodology

Defining the Problem

Machine learning models require training samples, each paired with an outcome value (class label) in order to learn how the data maps to the outcome classes. One training sample is made up of a set of features/variables and a class label, denoting which category group the sample belongs to (e.g., injured or non-injured)[see table 9.1]. The process of feature selection describes taking all the collected data, and from it identifying and extracting meaningful measures to describe the data.

The task to train a machine learning model to predict future injury was defined in two ways: **'a'** on

Table 9.1: A simple example of a training set with only 4 samples. Each sample has a set of features and a class label associated with it. The class label in classification tasks typically has an integer value which maps to some category i.e, 0 = non-injured group, and 1 = injured group.

	Feature_1	Feature_2	Feature_1	Classlabel
Sample 1	40.0	-3.12	23.60	0
Sample 2	22.0	-4.78	44.21	1
Sample 3	35.0	-3.99	39.02	0

a one sample per subject level, and **'b'** on a one sample per stride level. This means that on a sample per subject level, the extracted features described the subject's full data signal via a representative aggregated stride. Each subject is represented by a single sample in the training set. Whereas, on a sample per stride level the extracted features described a single stride of that subject's data only. Each subject is represented by a number of samples (equal to the number of strides in their data) in the training set.

These two different definitions of the problem pose distinctly different research questions, **'a'** can RRIs be predicted on a person level and **'b'** can RRIs be predicted on a single stride level.

9.2.2.1 Data Collection

Three hundred and eleven (311) participants were recruited to participate in a prospective study evaluating RRIs. This was the number of runners who responded to advertisements for volunteers for the study during the recruitment stage and subsequently attended an appointment for data collection. Full details of the data collection are provided in section 6.4. Data collection were carried out using both inertial sensors and Vicon retro-reflective markers. Inertial sensors were placed on the sacrum, and (left and right) thigh, tibia and foot of each participant to record accelerometer and gyroscope data. Vicon retro-reflective markers were placed on the participant following a custom Plug in Gait model [361], which captured the lower body, trunk and pelvis. Of this group, 150 participant's data were retained in order to build the injury predictive model. The reduction of the dataset was due to a number of human and data related issues. Thirty-five participants were withdrawn from the study for a number of reasons: health issues, pregnancy, non RRI or took up a new sport during the injury observation period. Seventeen eligible participants were removed for not completing the study in its entirety i.e., not taking part in the injury follow-up after the 1-year injury observation period and not providing all relevant clinical data. Data issues and incomplete data were the cause of further removal

of 109 participants for reasons such as: sensor failure, sensor configuration or software failure, data syncronisation issues, and incorrect sensor positioning during the course of data capture. As data were collected across 7 body segments, sometimes sensor errors caused the data from one segment to be lost. Only participants where complete segmental data were available were included in the analysis. The collected accelerometer and gyroscope data allowed the inclusion of loading and technique-based variables (features) to be defined and used to train the machine learning models. The Vicon data were used to extract features based on the collected running technique/kinematics.

Clinical data were collected on each participant with respect to: age, sex, weight, height, body mass index, running experience (novice or non-novice), training volume (running pace and running distance per week) and previous injury (yes or no) and recent injury (yes or no) [see table 9.2]. Participants took part in a one-time collection of their loading and technique data while treadmill running. Participants were tracked prospectively for 12-months and any participant who suffered a RRI in that time were assigned an 'injured' label in the dataset. A RRI was defined by Yamato et al. as *"any (training or competition) musculoskeletal pain in the lower limbs or lower back that causes a restriction/stoppage of running (distance, speed, duration, or training) for at least 7 days or 3 consecutive scheduled training sessions, or that requires the runner to consult a physician or other health professional"* [418]. Injury tracking involved contacting participants every 4-6 weeks to enquire about their current injury status. If an injury was reported, the participant was invited to have the injury inspected by two members of the research team who were qualified as a Chartered Physiotherapist and certified Athletic Trainer. If a participant could not attend this session, information regarding their injury complaints and any additional information given by their healthcare professional were recorded via phone call.

9.2.2.2 Data Screening and Preparation

Due to the large-scale size of the data collection, there was a large amount of data wrangling required. Inertial sensor data were downloaded from the sensor SD cards using Consensys software v1.4 provided by the inertial sensor manufacturer (Shimmer Sensing, Dublin, Ireland)). Inertial data were screened in an extended process in order to remove any erroneous data captured. In most cases there was no effective way to correct the errors (record failure, faulty configuration, sensor placement issues) and so these participants were removed as their full collected data was incomplete (i.e. missing left thigh). In order to build an accurate prediction classifier, we decided to only include complete, accurate representations of the data, to remove the variable of unknown error due to data manipulation or

imputation.

Vicon retro-reflective marker data were re-constructed and screened using a proprietary software Vicon Nexus (Vicon Oxford Metrics, Oxford, United Kingdom). Any missing markers in the optical motion analysis data were filled using an in-built spline-based gap filling with a 4th order low-pass Butterworth filter and cut off frequency of 15Hz. Vicon data were analysed stride-by-stride, and frames numbers were retained for a number of key events in each stride for each participant: initial contact (IC), peak-knee flexion (PKF) and toe-off (TO). This allowed the later extraction of joint angle values at each of these key events.

9.2.2.3 Data Segmentation

It is important to remember there are three types of data in the dataset collected: inertial sensor data (accelerometer and gyroscope), kinematic data (joint angles) and clinical data (age, sex, previous injury etc.). In order to align the inertial sensor data and kinematic data for an individual participant, cross correlation was performed using thigh segment orientation data produced by the optical motion analysis system and thigh segment orientation produced by the inertial sensor data. Thigh Y-axis orientation was chosen for dataset synchronisation as it provided the most stable signal in the inertial thigh orientation data compared to the other axes, and thigh specifically was chosen due to the close physical alignment between local coordinate systems of the thigh inertial sensor and the thigh placed retro-reflective markers. Thigh segment orientation was calculated using the accelerometer and gyroscope data in combination with the Madgwick 6DOF filter from Section 7.3. The inertial data were down-sampled to 200Hz to match the sampling rate of the kinematic data prior to performing the cross correlation. Due to the inaccuracies of 6-DOF orientation calculation described in section 7.3, some participants data could not be accurately aligned with the Vicon data. These participants had to be removed from the study as there was no other way to synchronise the inertial and kinematic systems of data.

Once the inertial based thigh orientation was aligned with the Vicon based thigh orientation, the timestamps of the first and last samples were stored, which represented the cropping region for this participant's captured data (for the segmentation step). For the subsequent machine learning-based analysis to predict injury, orientation in the form of joint angles from the optical camera-based motion analysis system were used instead of the inertial calculated orientation (as used in Chapters 7 and 8) due to its better stability and accuracy during running. From this point, reference to the orientation

will be expressed as joint angles or kinematic data.

With the cropping points stored, the inertial data for each participant were segmented into a single window representing their full running data capture (1-minute duration). The data of the 150 participants were stored for model training and testing. Instead of implementing a single train-test split of the data into two separate datasets, a Nested Cross validation was carried out in ordered to perform multiple train-test splits based on the 150 participant data [419, 420]. This will be described in further detail in sub-section 9.2.2.9 which describes the model training.

Table 9.2: Clinical patient data collected from each participant.

Clinical Data
Sex
Age
Weight (kg)
Height (m)
Body Mass Index
Training Load 1 - Distance per week
Training Load 2 - Running Pace
Running Experience[a]
Never Injured
Recently Injured[b]

a: Novice or non-novice runner, *b*: Recently Injured within 1-year: yes or no.

9.2.2.4   Feature Extraction

Features are described as either inertial features, kinematic features or clinical data features. Inertial features were extracted from the cropped region described in the Data Segmentation section (section 9.2.2.3). Any missing inertial data samples in the cropped section were filled with a linear interpolation in order to keep stride lengths aligned with the kinematic data. Each individual axis of a subject's inertial data (accelerometer and gyroscope X, Y, and Z axes on all segments) were split into windows capturing a single stride (approximately 80 strides per person). Inertial features were extracted from each X, Y and Z axis of the accelerometer and gyroscope data for each body segment (left and right thigh, tibia, foot and sacrum).

Kinematic features were extracted as joint angles from the Vicon kinematic data. Key events occurred during each stride for each participant: initial contact (IC), peak-knee flexion (PKF) and toe-off (TO). Joint angle values were extracted at each of these key events (IC, PKF and TO) for each stride, on each axis (X, Y, Z for all angles). Additionally maximum, minimum and range of motion

(ROM excursion) were extracted from each joint angle axis for each stride. The kinematic joints included in this experiment were: thorax and pelvis angles, and (left and right) hip, knee and ankle angles. This created a set of kinematic data to be used to train the machine learning models. These were collected for the left and right sides of the body, for each individual running stride. Clinical data features were collected via a subject survey.

Approach 'A': One sample per Subject

Following approach 'a', initial stride level features were extracted. From each of the approximately 80 stride windows for each runner, a small number of discrete values (representing inertial features) were extracted [maximum value, minimum value, rate of acceleration, max-min difference, time between max-mix peaks]. The choice of features were informed by RRI research, as these variables have been thought to potentially relate directly to injury (see discussion of loading and injury in Section 2.2.3.1). The maximum value represents the impact peak in the inertial data (measure of loading related to impact peak in vGRFs), and the rate of acceleration represents the rate of loading in the inertial data (related to rate of loading in vGRFs). In conjunction with these features, other inertial features from the frequency domain were extracted: the maximum and mean frequency per stride, and the resultant acceleration value across X, Y, Z axes.
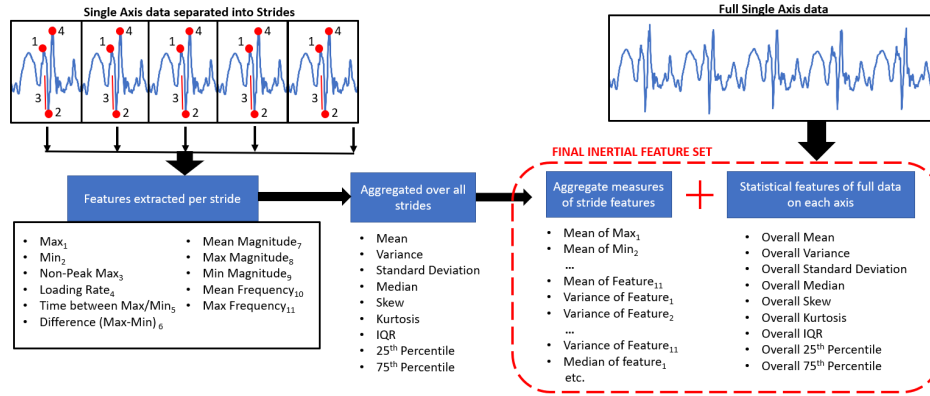
Figure 9.1: Feature extraction methodology for the inertial data features for a single axis of the accelerometer. This feature extraction process was repeated on all axes (X,Y, Z) of both the accelerometer and gyroscope sensors.

These discrete feature values were then aggregated (approach 'a') across all strides for that single subject axis (e.g. accelerometer X-axis of the sacrum) to represent a *single aggregated measure of that feature*. This transforms features from a stride-level measure to a subject-level measure (across all strides). Aggregation was carried out based on the mean, median, standard deviation, variance,

kurtosis, skew, interquartile range, 25th percentile and 75th percentile of the discrete values across all strides for each feature (see figure 9.1 for clarity). A range of aggregation strategies were included as it was not known whether certain transformations of the single-stride data to subject-level data were more informative in predicting injury. In addition to those features, subject-level statistical features were extracted. These were a fixed number of time domain features (mean, median, standard deviation, variance, kurtosis, skew, inter-quartile range, 25th percentile and 75th percentile) extracted from the full signals (80 strides combined) following a standard set of features used in time-series classification with inertial data [311, 334].

For approach 'a' the kinematic features were extracted in a similar way; within each of the approximately 80 strides, the magnitude of joint angles were extracted for: toe-off, initial contact, peak knee flexion, maximum value, minimum value and range of motion (excursion). These events were extracted for each joint angle: thorax, pelvis, hip, knee and ankle, for the left and right sides of the body, on all axes (X, Y, Z). These features were aggregated (approach 'a') across all strides for that single subject, to represent a single measure of each feature. Aggregation was carried out based on the mean, median, standard deviation, variance, kurtosis, skew, interquartile range, 25th percentile and 75th percentile of each feature across all strides of a particular runner.

Approach 'B': One Sample per Stride

Following approach 'b' stride level inertial features were extracted. This is initially similar to approach 'a' where a set number of features were extracted from each stride [maximum value, minimum value, rate of acceleration, max-min difference, time between max-mix peaks]. In addition, unique to approach 'b', a number of time-domain features (mean, median, standard deviation, variance, kurtosis, skew, interquartile range, 25th percentile and 75th percentile) and frequency domain (mean frequency, max frequency) features were extracted from *each stride* [rather than full data as in approach 'a']. This set of features are commonly used in time-series classification with inertial data [311, 334].

For the kinematic features in approach 'b', with each stride joint angles were extracted for: toe-off, initial contact, peak knee flexion, maximum value, minimum value and range of motion (excursion). These events were extracted for each joint angle: thorax, pelvis, hip, knee and ankle, for the left and right sides of the body, on all axes (X, Y, Z).

9.2.2.5   Difference between Approach 'A' and Approach 'B'

For approach **'a'** the set of inertial features used were the time domain features across the full subject signal (i.e. the standard deviation of the whole X-axis accelerometer signal for that subject) and an aggregation of discrete features extracted at a stride level (i.e. the mean maximum peak value across 80 strides). For approach **'b'** the set of inertial features used were the time and frequency domain features extracted per stride and the set of discrete features extracted per stride. The main difference in inertial and kinematic feature extraction for approach 'a' and 'b' is that in 'a', the aggregation performed transforms a stride-level feature to an aggregated feature which describes a given feature at a subject level. Whereas in 'b', there was no aggregation of features, each feature describes a measure at a stride level.

This distinction between 'a' and 'b' asks the question whether the likelihood of injury can be identified at a full person level, or a stride level; as in approach 'a' each subject represents one sample labelled injured or non-injured, versus 'b' where each stride represents one sample labelled injured or non-injured.

Separating the problem in this way leads to building two model types: one with 150 samples (each subject represents one sample) [Approach 'a'], and another with approximately 12000 training samples (each subject represents approximately 80 (stride) samples) [Approach 'b']. Traditionally in machine learning, more samples in training the model should increase its learning and predictive ability, and so stride-based samples may remove the need for more subject data. However, describing the problem in this way denotes predicting injury based on a single running stride alone which may not be possible [as in approach 'b']. It may be more beneficial to use an aggregated representation of a subjects data [as in approach 'a'], as this is both practically sensible in terms of the implications of an accurate predictive model, which would allow a individual runner's RRI risk to be assessed based on their full running data. Additionally, aggregation may also help mask any irregularities which exist on a stride-to-stride level.

9.2.2.6   Initial Statistical Analysis

In the first instance statistical univariate tests were performed on the features of the training data to assess whether a relationship existed between the features and injury, using Mann-Whitney U tests (continuous variables) and Chi-squared tests (categorical variables) in SPSS. These statistical tests were performed separately on the inertial, kinematic and clinical data features. Analyses were

performed twice, once for approach 'a' subject level features, and again for approach 'b' stride level features. The difference in the definition of the features between approach 'a' and 'b' may have an impact on the usefulness of the feature and its relationship with injury.

The decision was made to prioritise the analysis of effect sizes (Cohen's d) over p-values as when the sample size increases p-value significance testing is more likely to find a statistically significant relationship between variables. Effect size is uncoupled from sample size, and represents the magnitude of a relationship if one exists [421]. This can clarify whether any relationships are practically significant as well as statistically significant. Cohen's d values were interpreted as described in Cohen's work [422], where approximately d=0.2 represents a small effect size, d=0.5 represents a moderate effect size and d=0.8 represents a large effect size. The initial univariate statistical analyses of the relationship between the features and likelihood of injury were carried out using IBM SPSS Statistics 25.

9.2.2.7 Feature Selection

Due to the vast number of features, to avoid over-fitting and the curse of dimensionality, feature selection was carried out using *select-K-Best in Scikit-learn* [423]. This approach performs ANOVA tests between the features and injury labels and ranks features in order of f-scores, keeping the top K best scoring features. Feature selection was implemented in two ways: firstly, to retain only the top 20 features to train models (fixed features) [424] using *SelectKbest* and secondly, to include the 'optimal number of features' as a specific hyper-parameter to optimise during model training process (tuned features).

For a given classifier and given dataset (i.e. Inertial and Kinematic data) models were trained on; (i) basic classifiers with default parameters using the top 20 best features (from *SelectKbest*), (ii) classifiers with nested cross validation to tune best hyperparameters using the top 20 best features (from *SelectKbest*)[424], and (iii) classifiers with nested cross validation to tune both best hyperparameters and best feature set (including number of features as a hyperparameter). For brevity (i) and (ii) is included in the Appendix E Section 11.5 and Appendix F Section 11.6, and (iii) is described in the body of this chapter. A pipeline implementation was used involving feature selection (*SelectKBest* in Scikit-learn [423]), feature scaling (*StandardScaler* in sklearn [423]) and model creation (chosen machine learning classifier i.e. Random Forest). Feature scaling was included in the pipeline to ensure all data (features) were normalised for training and testing. The *StandardScaler* standardises features by removing the mean and scaling features to unit variance.

This is important as unscaled data may badly affect classifiers which are sensitive to the relative scale of feature, and also scaling features may lead to faster training times.

9.2.2.8   Hyperparameter Tuning

Hyperparameters are fixed values within each classifier which determine how the model is defined (e.g., 'K number of neighbours or the 'distance metric' used in K-Nearest Neighbour (KNN) [425]). The models were trained with hyper-parameter tuning via cross validation. Cross validation is a procedure by which the training data is split into a fixed number K of sub-sets called folds (K-folds). For K repetitions, one fold is selected to be test (validation) data, and the remaining K-1 folds are used for model training, until each fold has been selected once to evaluate the validation error of the trained model. Once all K folds have been used to test the model, the average validation accuracy is obtained via the average of the K validation error scores (one for each fold). Hyperparameter tuning with cross validation is done by defining a list of different values to try in the set of hyperparameters for a given classifier model. For each unique combination of hyperparameters, a model is trained with K-fold cross validation and tested on each fold. As described in the previous section, 'number of features' was included as a hyperparameter, along with the classifier defining hyperparameters (i.e., K number of neighbours in KNN). A single model with the set of hyperparameters which obtains the highest validation accuracy from the cross validation is selected as the best model (i.e. 10 features with K=4 nearest neighbours may have a higher validation score than 20 features with K=4 nearest neighbours or 20 features with K=3 nearest neighbours). This best model is re-trained with all the available training data. Cross validation is a tool used to help reduce the risk of overfitting the model during the training process.

Due to the large number of grid search combinations available between all hyperparameter sets, a randomised search approach was used (*RandomisedSearchCV* in Scikit-learn [423]), where it performed 1000 iterations of different hyperparameter combinations. This was a less exhaustive approach compared to grid search where every unique parameter combination is evaluated, but 1000 iteration was still suitable to search a large portion of the possible parameter space with lower computational overhead.

9.2.2.9   Model Training

In order to perform model training, evaluation and hyperparameter tuning a nested cross validation procedure was carried out. Nested cross validation involves two nested cross validation procedures. An outer cross validation (CV) performs the splitting of the full dataset into a training and test set using a 5-fold cross validation (K=5) [277, 278]. On the first iteration of the Outer CV (5-fold CV) the first *training* and *test* sets are created with 80% of participants in the training set and the remaining 20% of participants in the independent test set [274, 426]. If a participant's data were a member of the training set, then all of that subjects samples (one sample for approach 'a', many samples for approach 'b') were included in the training set only, and vice versa. This was an important step to induce randomisation while following good data hygiene ensuring that the data from any participant did not belong to both the training and testing set, which would bias the results. Following this spilt into training and testing, the Inner CV uses the *training set* split from the outer CV and performs a 10-fold CV in the Inner loop on this training data only in order to perform parameter tuning (described in the previous section). 10-fold cross validation is believed to be one of the most robust implementations of K-fold and has become a standard way of predicting the error rate, especially when there is limited training data [427]. Parameter tuning leads to an optimal model (highest average validation score), which is passed back to the Outer CV and evaluated using the *test* set which was kept independent and split from the training data. Evaluating the best model will produce an accuracy score which was stored, along with the hyperparameters which defined this best model. This cycle represents one iteration of the Outer CV loop and is repeated 5 times in total, producing 5 best models from each Inner CV loop with best hyperparameters and scores. Averaging the best model scores (scored with AUC) across the 5 best models provides a reliable estimation of the generalisation test error of the given model building procedure [428–430]. This final testing score explains how well the model can generalise to unseen data and how well it may perform in a clinical setting. This estimation of the test error is thought to be less biased as a number of splits of the data were made to train/test the models, keeping test data separate until calculating the generalised model error [419]. A single-split implementation may suffer from an overly optimistic/or pessimistic estimation of the generalisation error due to the splits of the training and test data [429]. It is important that the distribution of the test data is similar to the distribution of the training data (e.g. different distributions such as training the model using data of athletes and testing the model using data of novice runners).

There does not seem to be a clear single approach followed using nested cross validation in previous

research to produce a 'final' model for future testing and deployment purposes. Three main approaches have been identified. Firstly, if the 5 best models are stable and have similar hyperparameter values, one of these models can be selected and trained on the entire training data after the nested CV procedure [431]. If the 5 best models have vastly different best parameter values, then this approach may not be appropriate as these parameters were best for simply one split of the whole dataset (one outer iteration and test split). This may be an issue if that split is not representative of the true distribution of all test cases [429]. Secondly, an ensemble model may be created using each model from the outer loop which can use majority voting to assign classes to new unseen data [432]. Thirdly, the procedure of the Inner CV (10-fold CV in this case) is applied one final time to the entire dataset in order to tune hyperparameters to the full dataset, employing the model procedure (inner loop) on all data [430, 433]. This approach seems the most appropriate as the model building procedure (specific classifier + cross validation) is what is being compared, and so the best model building procedure can then be applied to all data. The generalisation error of these models is simply the average error initially calculated using the Nested CV [428] or of course if there are additional data available, these can be used to test the true model accuracy [430]. The Nested cross validation approach validates the estimation of the generalisation error of the model building procedure itself (training, validation, tuning and testing) and not necessarily a particular model [420, 429].

The training data was used to create six datasets: inertial data only (I), clinical data only (C), kinematic data only (K) and combinations of: I-C, I-K and I-C-K. Each different machine learning classifier was trained on each of the six datasets to see which classifiers performed better on different data sources. Seven common traditional machine learning classifiers were implemented: K-Nearest Neighbour, Naive Bayes, Decision Trees, Logistic Regression, Random Forest, Multi-Layer Perceptron and an Ensemble of Support Vector Machine Bagging classifier (Bagged SVM ensemble). Each of these approaches are explained in Chapter 5 Section 5.2.1. These classifiers were all implemented in Python 3.7 using Scikit-learn built-in methods [423]. The chosen classifiers cover a range of approaches which have previously been used for classification/prediction using inertial data [e.g., injury risk prediction (section 5.3.1), disease prediction (section 5.3.2), activity recognition (section 5.3.3)]. Some classifiers may be better suited at identifying the relationships within the data for the problem of injury prediction.

9.2.3 Results

The results from this analysis were not used to influence or guide any following analyses, this was performed solely to evaluate whether there was a relationship between the features and injury. The top 20 resulting effect sizes of the features can be seen in ranked descending order in tables 9.3 and 9.4. The features were not normally distributed and so Mann-Whitney U tests were used to examine the continuous variables using Cohen's d effect sizes. Following the interpretation of Cohen's d in Cohen [422], it can be seen that the majority of features in our study have a weak effect size.

Initial Statistics Approach 'A'

As can be seen from the results mostly weak relationships existed between the features and injury for approach 'a'. Of the 5412 features total, 4 features [Sex, Never Injured, Recently Injured and Running Experience] were evaluated with Chi Square tests which resulted in low effect sizes (with Phi score magnitudes of 0.01 to 0.21, which represented weak-small effects [434]). Of the remaining 5408 features, when the results of the Mann-Whitney U tests were sorted from largest absolute effect size (ignoring direction of effect) the top 20 features ranged in Cohen's D from 0.22 to 0.31 which represent small effect sizes [422] (see table 9.3 for the top 20 features as selected by Mann-Whitney U tests.).

Table 9.3: This table shows the results of the Mann-Whitney U tests when sorted from largest absolute effect size (ignoring direction of effect) and selecting the top 20 features with the largest effect size (Approach A).

Feature	Z	p	Effect Size	Absolute Effect size
Timebetw_gyroY_Sacrum_skew	-3.82	0	-0.31	0.31
ThoraxAngles_fle_ROM_R_kurt	-3.20	0.001	-0.26	0.26
Lfoot_globGYRY_median	-3.16	0.002	-0.26	0.26
MaxfreqR_Accz_LpSha_kurt	-3.05	0.002	-0.25	0.25
MeanfreqR_Accy_LpSha_iqr	-2.94	0.003	-0.24	0.24
AvgMin_Accy_LpSha_kurt	-2.91	0.004	-0.24	0.24
MeanfreqR_Accy_LpSha_var	-2.88	0.004	-0.24	0.24
MeanfreqR_Accy_LpSha_std	-2.88	0.004	-0.24	0.24
Diffavg_gyroZ_RpThi_iqr	-2.85	0.004	-0.23	0.23
AvgMin_gyroY_LpThi_mean	-2.82	0.005	-0.23	0.23
Diffavg_gyroY_LpThi_mean	-2.80	0.005	-0.23	0.23
MeanMag_gyroX_LpFoo_kurt	-2.73	0.006	-0.22	0.22
Sacrum_globACCZ_skew	-2.68	0.007	-0.22	0.22
KneeAngles_abd_IC_L_mean	-2.67	0.007	-0.22	0.22
KneeAngles_abd_IC_L_per25	-2.67	0.008	-0.22	0.22
Diffavg_gyroY_Sacru_skew	-2.67	0.008	-0.22	0.22
AvgMax_gyroZ_RpSha_skew	-2.66	0.008	-0.22	0.22
Lfoot_globGYRY_skew	-2.66	0.008	-0.22	0.22
AvgMin_gyroZ_RpThi_kurt	-2.65	0.008	-0.22	0.22
AvgMax_gyroY_Sacru_skew	-2.64	0.008	-0.22	0.22

Initial Statistics Approach 'B'

Of the 949 features total, 4 categorical features [Sex, Never Injured, Recently Injured and Running Experience] were evaluated with Chi Square tests which resulted in low effect sizes (with Phi score magnitudes of 0.01 to 0.22, which represented weak-small effect sizes [434]). Of the remaining 945 features, when the results of the Mann-Whitney U tests were sorted from largest absolute effect size (ignoring direction of effect) the top 20 features ranged in Cohen's D from 0.17 to 0.23 which represent small effect sizes [422] (see table 9.4 for the top 20 features as selected by Mann-Whitney U tests).

Table 9.4: This table shows the results of the Mann-Whitney U tests when sorted from largest absolute effect size (ignoring direction of effect) and selecting the top 20 features with the largest effect size (Approach B).

Feature	Z	p	Effect Size	Absolute Effect size
KneeAngles_abd_IC_L	-24.93	<0.0001	-0.23	0.23
median_gyroY_L_Foot	-23.95	<0.0001	-0.22	0.22
KneeAngles_abd_IC_R	-23.91	<0.0001	-0.22	0.22
KneeAngles_abd_MAX_L	-22.84	<0.0001	-0.21	0.21
KneeAngles_abd_MAX_R	-22.84	<0.0001	-0.21	0.21
Skew_Accz_Sacrum	-22.60	<0.0001	-0.21	0.21
Skew_gyroY_L_Foo	-22.37	<0.0001	-0.20	0.20
Mean_Accz_R_Shank	-21.76	<0.0001	-0.20	0.20
KneeAngles_abd_TO_R	-21.28	<0.0001	-0.19	0.19
KneeAngles_abd_TO_L	-20.91	<0.0001	-0.19	0.19
Kurtosis_gyroX_Sacrum	-20.44	<0.0001	-0.19	0.19
KneeAngles_abd_MIN_L	-20.27	<0.0001	-0.18	0.18
Per75_Accy_R_Foot	-19.57	<0.0001	-0.18	0.18
ThoraxAngles_abd_IC_L	-19.36	<0.0001	-0.18	0.18
IQR_Accy_R_Foot	-19.33	<0.0001	-0.18	0.18
Timebetw_gyroX_R_Shank	-18.88	<0.0001	-0.17	0.17
KneeAngles_abd_PKF_L	-18.85	<0.0001	-0.17	0.17
Loadingrate_Accz_R_Thigh	-18.42	<0.0001	-0.17	0.17
Minimum_gyroX_Sacrum	-18.39	<0.0001	-0.17	0.17
Median_Accy_L_Thigh	-18.30	<0.0001	-0.17	0.17

Model Accuracy

These results point to a weak relationship between the representation of the features and injury, which may lead to poor injury prediction when analysed at a univariate level. However, machine learning allows for the possible identification of implicit patterns in the data that may be related to injury prediction, which may not be identified by univariate statistics.

For each of the seven classifiers [K-Nearest Neighbour, Naive Bayes, Decision Trees, Logistic Regression, Random Forest, Multi-Layer Perceptron and Bagged SVM ensemble], models were separately trained with each combination of: inertial data only (I), clinical data only (C), kinematic data only (K) and combinations of: IC, IK and ICK.

For a given classifier, and given data set (i.e. IK) models were trained on; (i) basic classifiers

with default parameters using the top 20 best features, (ii) classifiers with nested cross validation to tune best hyperparameters using the top 20 best features and (iii) classifiers with nested cross validation to tune both best hyperparameters and best feature set (including number of features as a hyperparameter). Due to the similarity of results across these approaches, for brevity the results for (i) and (ii) are included in the Appendix (Appendix E Section 11.5) and Appendix F Section 11.6) and the results for (iii) are outlined in this section. The performance of the machine learning models were evaluated using the ROC-AUC scoring metric. This was chosen as it is less sensitive to class imbalances compared to accuracy scoring evaluated on the true classification rate alone [280]. Precision, recall and F1 scores were evaluated following the formulae outlined in equations 9.1, 9.2 and 9.3. The scores for each model of (iii) can be seen in tables of the appendix section 11.4.

$$\text{Accuracy } = \frac{\text{TN} + \text{TP}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{9.1}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{9.2}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{9.3}$$

Results: Approach 'A'

The 150 subject training dataset contained the prospective data of 77 uninjured and 73 injured runners. On each of 5 iterations of the outer Nested Cross Validation loop, the dataset was split into 80% of samples for training and the remaining 20% of samples for an independent test set [274, 277]. For approach 'a', each subject was represented by one sample in either the training (120 samples) or testing (30 samples) sets.

With respect to results on average, the average testing accuracy of each classifier across all sets of data (I, C, K, IC, CK, ICK) ranged in AUC-ROC from 0.50-0.53, which overall is quite poor. The highest average of 0.53 was the result of the average AUC-ROC of all datasets (I, C, K, IC, CK, ICK) trained on Multi-Layer Perceptron. The average testing accuracy of each set of data across all classifiers [K-Nearest Neighbour, Naive Bayes, Decision Trees, Logistic Regression, Bayes Ridge Regression, Random Forest, Multi-Layer Perceptron and Bagged SVM ensemble] ranged in AUC-ROC from 0.46 - 0.56, which on the higher end is still a relatively weak classifier. The average across all

classifiers trained on clinical features (C: 0.56 ROC-AUC) had the highest average accuracy. Even though these were the best testing scores on average (across datasets and across classifiers), an ROC-AUC score of 0.53-0.56 is a measure of weak predictive ability. These average scores can be seen highlighted in blue in tables 11.1 and 11.2 in the Appendix D section 11.4.

Figure 9.2: Box-plots showing the mean AUC-ROC score of each classifier trained on the Clinical dataset for approach 'a'. The classifiers used were 'LogR': Logistic Regression, 'NB': Naive Bayes, 'KNN': K-Nearest Neighbour, 'DT': Decision Tree, 'RF': Random Forest, 'MLP': Multi-Layer Perceptron, 'BaggingSVM': Support Vector Machine Ensemble (bagging) classifier.

With regards to individual models, models with an AUC-ROC score greater than 0.55 were identified as the 'better models' and compared in terms of their AUC-ROC accuracy, precision, recall, F1-scores and confusion matrices. Seven of the overall 42 trained models [7 classifiers X 6 data sets (I, C, K, IC, CK, ICK) = 42 models] scored >0.55 ROC-AUC on average across the independent test sets within nested cross validation. Interestingly, the clinical feature set (C) was associated with the greatest number of 'better models' with 5/7 classifiers scoring an ROC-AUC >0.55 when trained using this data. The testing accuracy of the 7 models can be seen in table 9.5 where AUC-ROC score ranged from 0.55 - 0.58. Box-plots of each classifier trained on the Clinical dataset can be seen in figure 9.2. Examination of the AUC-ROC in combination with the additional measures (precision, recall, F1 score and confusion matrices) was necessary in order to gain a better understanding of how well the models performed.

The F1 score is the harmonic mean of the recall and precision, which means that it takes into account both false-positives and false-negatives. This is a helpful metric of model success when paired with the confusion matrices. It can be seen in table 9.6 where the model accuracies are ranked in terms of F1 score, and the associated confusion matrices in figure 9.3 that none of the models are particularly strong. The highest F1 score of 0.54 came from a KNN classifier trained on the C dataset (Clinical features).

Table 9.5: This tables shows the results of the AUC-ROC scores, precision, recall and F1 scores for the top 7 'best models' which had an AUC-ROC accuracy >0.55. The models are grouped in terms of dataset used to train the model in descending order of AUC-ROC score.

Dataset	Classifier	AUC-ROC Mean ± St.D	Precision	Recall	F1
C	NB	0.58 ± 0.08	0.45	0.63	0.52
C	RF	0.58 ± 0.10	0.60	0.50	0.53
C	LogR	0.56 ± 0.04	0.56	0.51	0.53
C	KNN	0.56 ± 0.03	0.55	0.53	0.54
C	DT	0.55 ± 0.06	0.54	0.51	0.52
I	MLP	0.57 ± 0.05	0.56	0.33	0.37
IC	MLP	0.55 ± 0.06	0.64	0.42	0.45

KNN: K-Nearest neighbour, RF: Random Forest, DT: Decision Trees, NB: Naive Bayes, LogR: Logistic Regression, MLP: Multi-Layer Perceptron.

Recall is an important metric given the objective of these models; to identify those potentially predisposed to injury, and refer them for an injury prevention intervention. It is arguable that in this case false negatives (classifying those likely to become injured as non-injured) are more important to prevent than false positives, as there is no harm associated with healthy runners undergoing injury prevention strategies. An Naive Bayes classifier trained on the C dataset had the best recall and true positive rate where 62% of injured samples were correctly identified. This model had a AUC-ROC of 0.58 ± 0.08, and precision, recall and F1 score of 0.45, 0.63 and 0.52, respectively.

Table 9.6: This tables shows the results of the AUC-ROC scores, precision, recall and F1 scores for the top 7 'best models' which had an AUC-ROC accuracy >0.55. The models are ranked in terms of F1 score in descending order.

Dataset	Classifier	AUC-ROC Mean ± St.D	Precision	Recall	F1
C	KNN	0.56 ± 0.03	0.55	0.53	0.54
C	LogR	0.56 ± 0.04	0.56	0.51	0.53
C	RF	0.58 ± 0.10	0.60	0.50	0.53
C	NB	0.58 ± 0.08	0.45	0.63	0.52
C	DT	0.55 ± 0.06	0.54	0.51	0.52
IC	MLP	0.55 ± 0.06	0.64	0.42	0.45
I	MLP	0.57 ± 0.05	0.56	0.33	0.37

KNN: K-Nearest neighbour, RF: Random Forest, DT: Decision Trees, NB: Naive Bayes, LogR: Logistic Regression, MLP: Multi-Layer Perceptron.

Although none of the models had strong mean predictive ability, the highest accuracy of possible
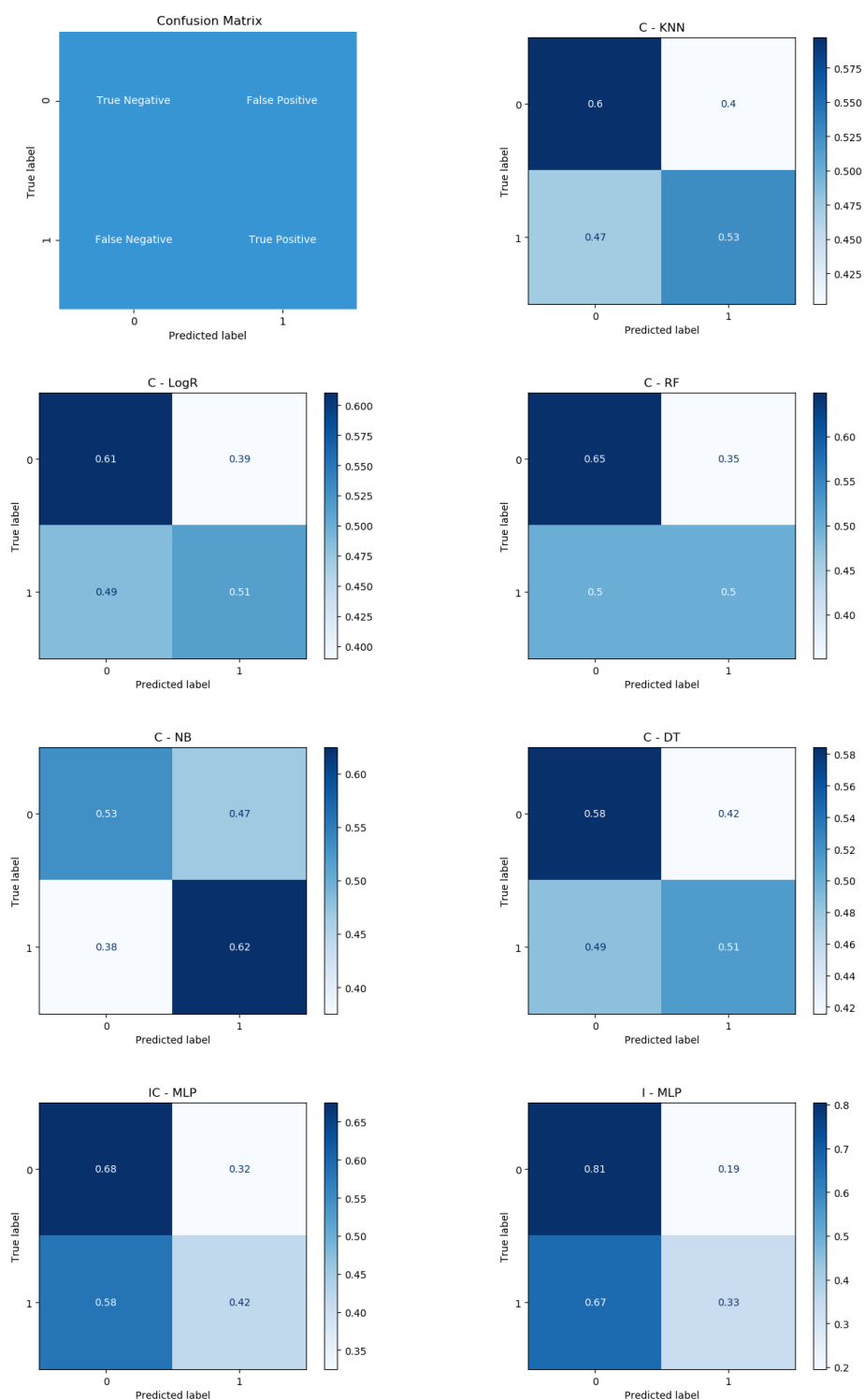
Figure 9.3: Table of normalised confusion matrices of the top 7 predictive models for Approach 'a', where 0: non-injured group, 1: injured group. Note: KNN: K-Nearest neighbour, RF: Random Forest, DT: Decision Trees, NB: Naive Bayes, LogR: Logistic Regression, MLP: Multi-Layer Perceptron.

scores (taking into account standard deviations) were 0.66 and 0.68 AUC using the Naive Bayes and Random Forest classifiers with the clinical dataset, respectively. These scores begin to move away from weak prediction (0.58 AUC) into classifiers with some predictive value. However, these higher accuracies are dependent on the split of the data and so a more stable model (smaller range of standard deviation) would be preferred if the model was to be considered for deployment. Even though the C-NB model (Naive Bayes trained on dataset C) had better recall for true positives, the range of possible model error (0.58 ± 0.08) is larger compared to the two models C-KNN and C-LogR which had a better F1 scores, and better spread of standard deviations of only 0.03 and 0.04, respectively. The plots contained in figure 9.5 display the ROC curves of the top two F1 score models in comparison to a guide to interpret the curves. It can be seen that the ROC curves for these models are close to the straight line curve representing results related to random chance.

Analyses were carried out where a random classifier was constructed and evaluated over 100000 times via a Monte Carlo simulation where the results of the random classifier were visualised as a histogram to establish the region of AUC-ROC scoring where a randomly higher or lower score was possible. This was particularly important if the best model accuracy fell within this region as we could have found a model with higher AUC-ROC due to chance. The random classifier was set up in a similar style to approach 'a' where the same data splits (of training and testing) from the nested CV were used. Therefore 5 random classifiers were implemented to match with the 5 split test sets and an average accuracy across the 5 random classifiers were saved for each 100000 random iterations. After 100000 random iterations it was found that the spread of AUC-ROC scores for the random classifiers fell in the region of 0.36 - 0.64 (see figure 9.4). This result shows that the score of 0.55- 0.58 AUC-ROC achieved by the best models were not better than a random predictor.

If the resulting models had a high level of injury prediction accuracy, it would be important to be able to identify any injury causative factors which helped the machine learning models to distinguish between the two groups. Using the top two models from figure 9.3, feature importance was determined by the number of times certain features were selected as the best hyperparameters across the 5 iterations of the Outer loop during Nested Cross Validation. The features which contributed the most to these two models can be seen in tables 9.7 and 9.8, for the K-Nearest Neighbour and Logistic Regression classifiers, respectively.
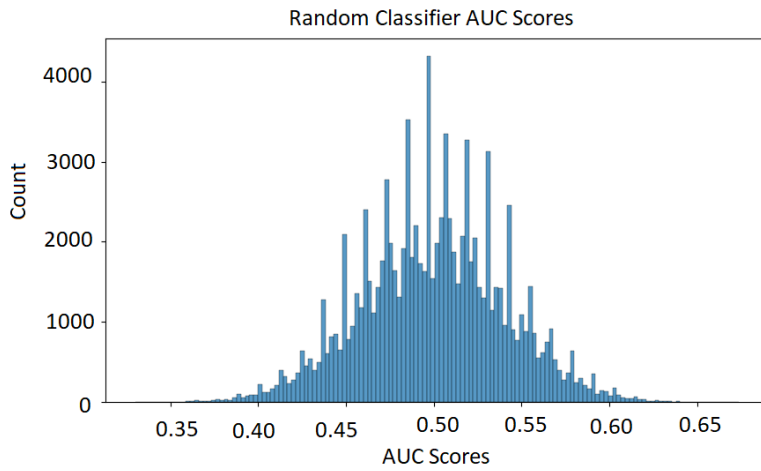
Figure 9.4: The spread of AUC-ROC scores produced by 100000 iterations of a random classifier for approach 'a'.

Table 9.7: The number of times each feature was selected for the 'Clinical feature trained K-Nearest Neighbour' as part of the hyperparameters search for the best 5 model iterations during Nested Cross Validation. The best 5 models were used to calculate the average accuracy of the K-Nearest Neighbour model (0.56 AUC-ROC), and the following features appeared a number of times across the best models.

Feature	Feature Importance
Running Experience[a]	5/5
Recently Injured[b]	5/5
Previously Injured[c]	5/5
Running Distance	3/5
Weight	3/5
Height	3/5
Sex	1/5
Body Mass Index	1/5
Running Pace	1/5

a: Novice or non-novice runner, *b*: Recently Injured within 1-year: yes or no, *c*: Previously Injured: yes or no.
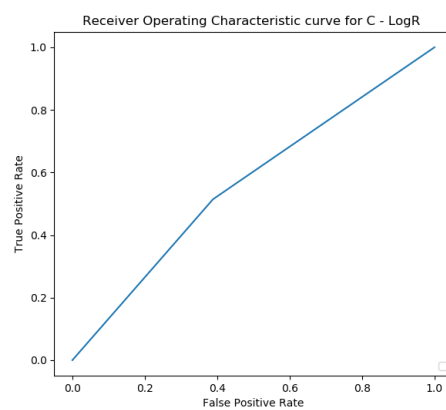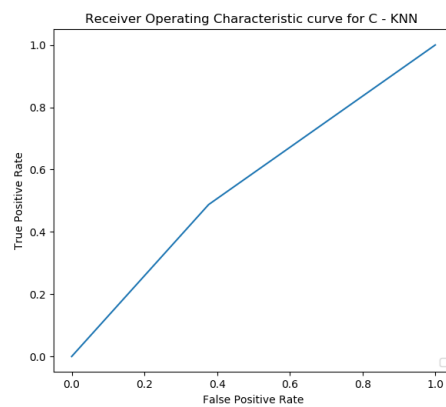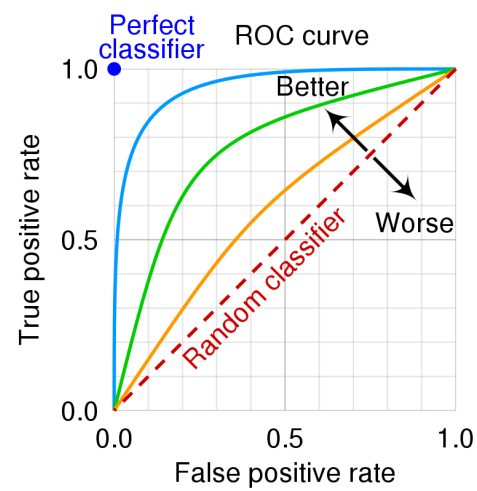
Figure 9.5: Receiver Operating Characteristic curves for the top two (F1 ranked) models in approach 'a' which show that the model performances with respect to the true positive rate and false positive rate are close to a random classifier.

Table 9.8: The number of times each feature was selected for the 'Clinical feature trained Logistic Regression model' as part of the hyperparameters search the best 5 model iterations during Nested Cross Validation. The best 5 models were used to calculate the average accuracy of the Logistic Regression model (0.56 AUC-ROC), and the following features appeared a number of times across the best models.

Feature	Feature Importance
Running Experience[a]	5/5
Recently Injured[b]	5/5
Previously Injured[c]	5/5
Running Distance	5/5
Weight	5/5
Height	5/5
Age	4/5
Body Mass Index	4/5
Sex	3/5
Running Pace	3/5

a: Novice or non-novice runner, *b*: Recently Injured within 1-year: yes or no, *c*: Previously Injured: yes or no.

Results: Approach 'B'

The 150 subject training dataset contained the prospective data of 77 uninjured and 73 injured runners. On each of 5 iterations of the outer Nested Cross Validation loop, the dataset was split into 80% of samples for training and the remaining 20% of samples for an independent test set [274, 277]. For approach 'b', each subject was represented by a number of running strides, where each single stride was one sample in either the training (approximately 9500 samples) or testing (approximately 2500 samples) sets.

With respect to results on average, the average testing accuracy of each classifier across all sets of data (I, C, K, IC, CK, ICK) ranged in AUC-ROC from 0.48-0.52, which is poor prediction accuracy. The highest average of 0.52 was the result of the average of all datasets (I, C, K, IC, CK, ICK) trained on Decision Tree classifiers, however this result is equivalent to chance and of no importance. The average testing accuracy of each set of data across all classifiers [K-Nearest Neighbour, Naive Bayes, Decision Trees, Logistic Regression, Random Forest, Multi-Layer Perceptron and Bagged SVM ensemble] ranged in AUC-ROC from 0.47 - 0.57, which on the higher end is still a relatively weak classifier. The average across all classifiers trained on clinical features the highest level of average accuracy (C: 0.57 ROC-AUC). Even though these were the best testing scores on average (across datasets and across classifiers), an ROC-AUC score of 0.52-0.57 shows poor to weak predictive ability.

These average scores can be seen highlighted in blue in tables 11.3 and 11.4 in the Appendix E section 11.4.
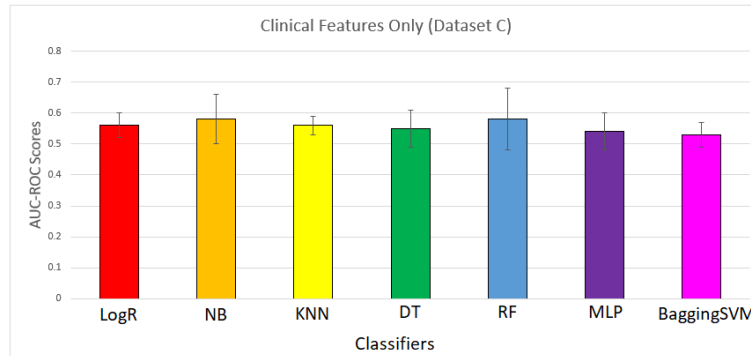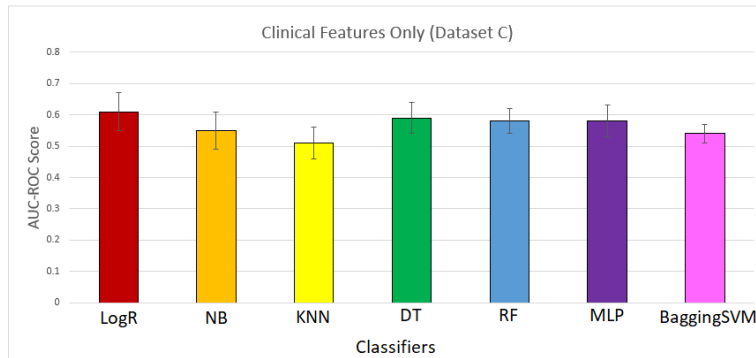

Figure 9.6: Box-plots showing the mean AUC-ROC score of each classifier trained on the Clinical dataset for approach 'b'. The classifiers used were 'LogR': Logistic Regression, 'NB': Naive Bayes, 'KNN': K-Nearest Neighbour, 'DT': Decision Tree, 'RF': Random Forest, 'MLP': Multi-Layer Perceptron, 'BaggingSVM': Support Vector Machine Ensemble (bagging) classifier.

With respect to individual models, models with an AUC-ROC score greater than 0.55 were identified as the 'better models' and compared in terms of their AUC-ROC accuracy, precision, recall, F1-scores and confusion matrices. 6 of the overall 42 trained models [7 classifiers X 6 data sets (I, C, K, IC, CK, ICK) = 42 models] scored >0.55 ROC-AUC on the independent test set. Similarly to the results for approach 'a', the clinical feature set (C) was associated with the most number of 'better models' with 5/6 classifiers scoring an ROC-AUC >0.55 when trained using this data. The testing accuracy of the 6 models can be seen in table 9.9 where AUC-ROC score ranged from 0.55 - 0.61. Box-plots of each classifier trained on the Clinical dataset can be seen in figure 9.6. Examination of the AUC-ROC in combination with the additional measures (precision, recall, F1 score and confusion matrices) was important to get a better understanding of how well the models performed.

With a focus on F1 scores, it can be seen in table 9.10 where the model accuracies are ranked in terms of F1 score, and the associated confusion matrices in figure 9.7 that the 'best models' are relatively weak classifiers. The highest F1 score of 0.61 came from a Logistic Regression classifier trained on the C dataset (Clinical features). This model had a AUC-ROC of 0.61 ± 0.06, and precision, recall and f1 score of 0.60, 0.61 and 0.61, respectively. Also taking recall into account, the MLP classifier trained on the C dataset had the highest recall (0.67), this was the second most accurate model in terms of AUC-ROC (0.58 ± 0.05). With the aim of this thesis in mind, primarily we wish to chose a model which maximises true positives, while minimising false negatives; we want to

Table 9.9: This tables shows the results of the AUC-ROC scores, precision, recall and F1 scores for the top 6 'best models' which had an AUC-ROC accuracy >0.55. The models are grouped in terms of dataset used to train the model in descending order of AUC-ROC score.

Dataset	Classifier	AUC-ROC Mean ± St.D	Precision	Recall	F1
C	LogR	0.61 ± 0.06	0.60	0.61	0.61
C	DT	0.59 ± 0.05	0.60	0.4	0.49
C	RF	0.58 ± 0.04	0.62	0.42	0.50
C	MLP	0.58 ± 0.05	0.57	0.67	0.60
C	NB	0.55 ± 0.06	0.43	0.60	0.50
K	KNN	0.56 ± 0.03	0.55	0.58	0.57

KNN: K-Nearest neighbour, RF: Random Forest DT: Decision Trees NB: Naive Bayes LogR: Logistic Regression MLP: Multi-Layer Perceptron.

identify as many runners as possible which may predisposed to injury. For the problem of identifying those predispose to injury, a certain portion of false positives is not harmful, as the outcome for this specific model would be a runner who is not likely to become injured would be directed to follow injury prevention strategies. However, it is clear we want to avoid models that simply classify all samples as 'injured', and so a trade-off between true negatives and true positives is required. Therefore, through consulting the F1 score, recall and confusion matrices, the most useful prediction model developed for approach 'b' was (as previously mentioned) the Logistic Regression classifier trained on the C dataset, and also the Multi-Layer Perceptron classifier which classifies a greater portion of injured samples correctly at the expense of more false positives which reduces the overall accuracy.

Table 9.10: This tables shows the results of the AUC-ROC scores, precision, recall and F1 scores for the top 6 'best models' which had an AUC-ROC accuracy >0.55. The models are ranked in terms of F1 score in descending order.

Dataset	Classifier	AUC-ROC Mean ± St.D	Precision	Recall	F1
C	LogR	0.61 ± 0.06	0.60	0.61	0.61
C	MLP	0.58 ± 0.05	0.57	0.67	0.60
K	KNN	0.56 ± 0.03	0.55	0.58	0.57
C	NB	0.55 ± 0.06	0.43	0.60	0.50
C	RF	0.58 ± 0.04	0.62	0.42	0.50
C	DT	0.59 ± 0.05	0.60	0.40	0.49

KNN: K-Nearest neighbour, RF: Random Forest DT: Decision Trees NB: Naive Bayes LogR: Logistic Regression MLP: Multi-Layer Perceptron.

The two best models C-KNN and C-LogR performed the best in terms of AUC-ROC accuracy and
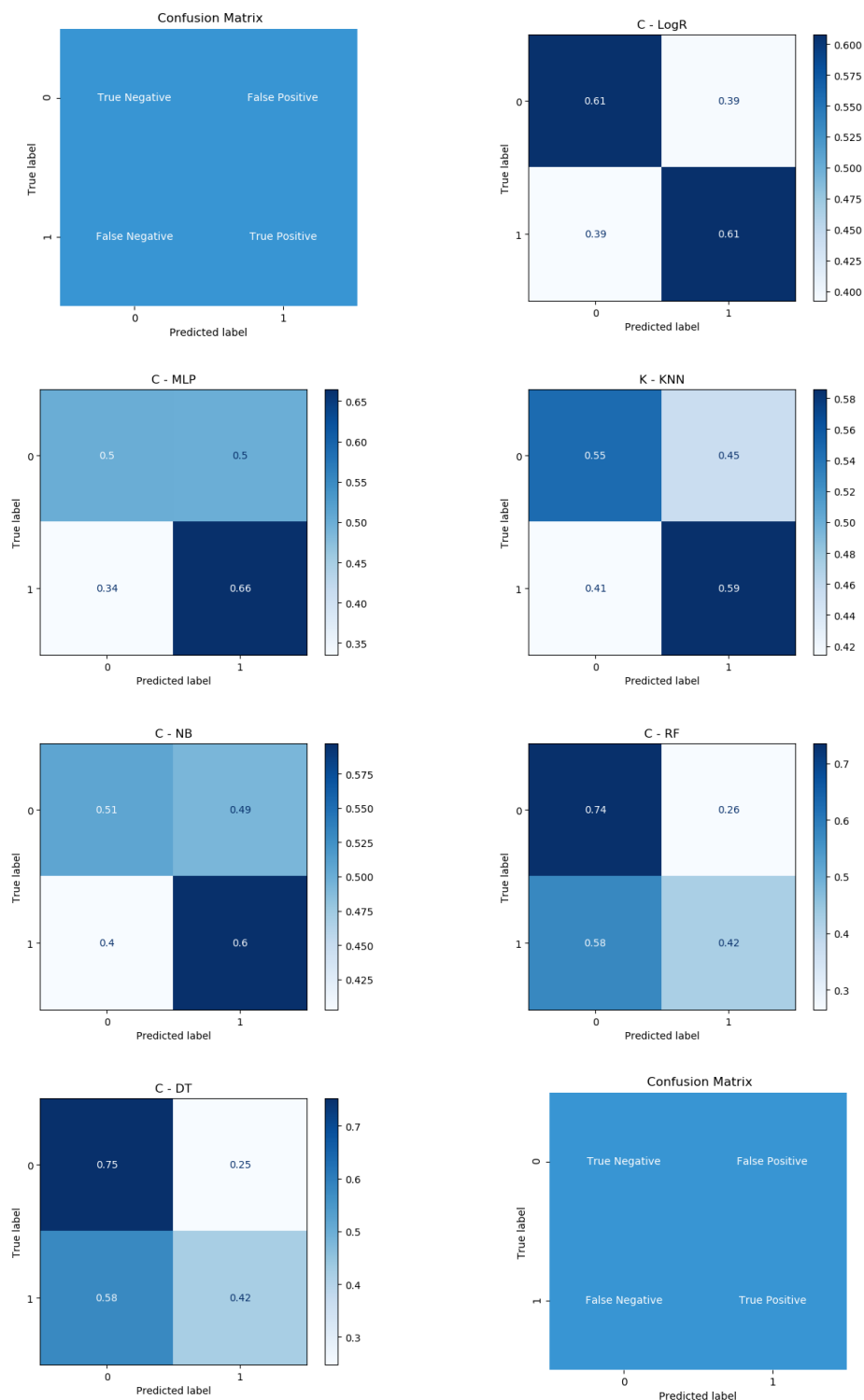
Figure 9.7: Table of normalised confusion matrices of the top 6 predictive models for Approach 'a', where 0: non-injured group, 1: injured group. Note: KNN: K-Nearest neighbour, RF: Random Forest DT: Decision Trees NB: Naive Bayes LogR: Logistic Regression MLP: Multi-Layer Perceptron.
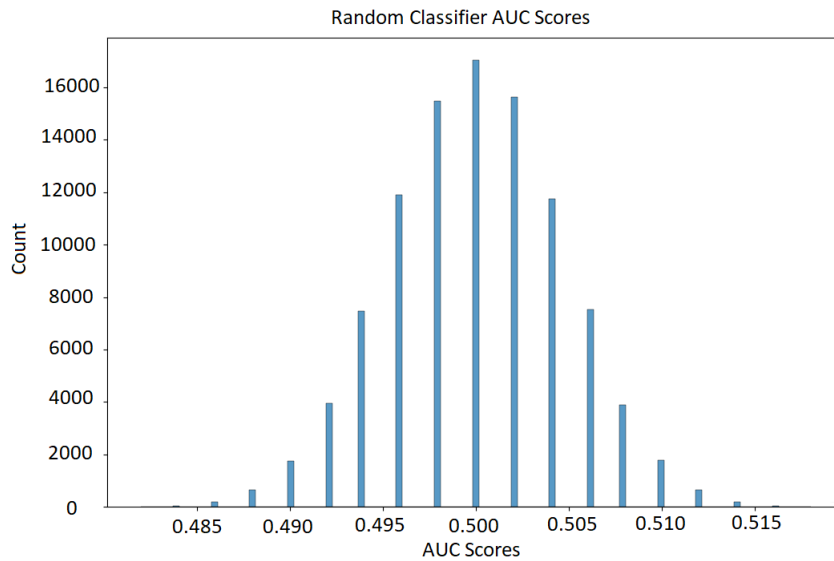
Figure 9.8: The spread of AUC-ROC scores produced by 100000 iterations of a random classifier for approach 'b'.

F1 score. The plots contained in figure 9.9 display the ROC curves of the top two models in comparison to a guide to interpret the curves. It can be seen that the ROC curves for these models are deviating away from the random classifier curve more than the two top models for approach 'a'. These models were shown to be better than random chance via the ROC curves and through implementing a random classifier for approach 'b'. A random classifier was constructed and evaluated over 100000 times via a Monte Carlo simulation where the results of the random classifier established a region of random chance related AUC-ROC scores. Five random classifiers were implemented to match with the 5 split test sets and average accuracy across the 5 random classifiers were stored for each 100000 random iterations. After 100000 random iterations it was found that the spread of AUC-ROC scores for the random classifiers fell in the region of 0.48 - 0.52 (see figure 9.8). This suggests that the scores of 0.55 - 0.61 AUC-ROC achieved by the best models for approach 'b' were better than random chance, and that some learning of the features occurred during training. The highest accuracy of possible scores (taking into account standard deviations) were 0.64 and 0.67 AUC using the Decision Tree and Logistic Regression classifiers with the clinical dataset, respectively (see table 9.9). These higher scores slowly move the model accuracy towards having some predictive value.

Using the top two models from table 9.10, feature importance was determined by the number of times certain features were selected as the best hyperparameters across the 5 iterations of the Outer loop during Nested Cross Validation. The features which contributed the most to these two models can be seen in tables 9.11 and 9.12, for the Logistic Regression and Multi-Layer Perception classifiers,
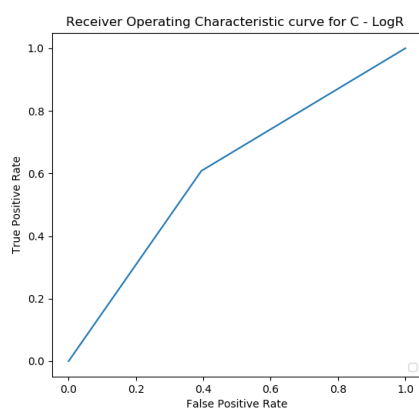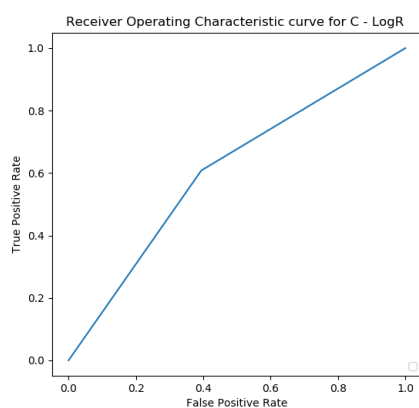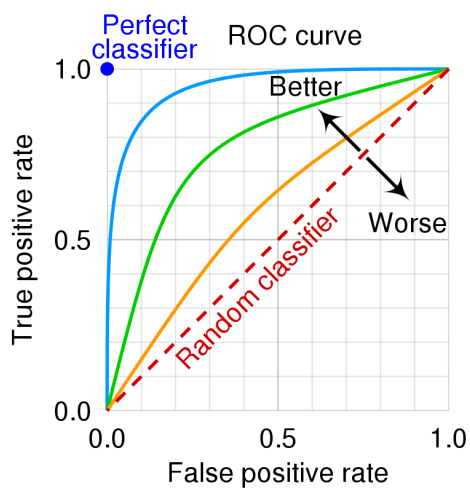
Figure 9.9: Receiver Operating Characteristic curves for the top two (F1 ranked) models in approach 'b'.

respectively.

Table 9.11: The number of times each feature was selected for the 'Clinical feature trained Logistic Regression model' as part of the hyperparameters search for the best 5 model iterations during Nested Cross Validation. The best 5 models were used to calculate the average accuracy of the Logistic Regression model (0.61 AUC-ROC), and the following features appeared a number of times across the best models.

Feature	Feature Importance
Running Experience[a]	5/5
Recently Injured[b]	5/5
Previously Injured[c]	5/5
Running Distance	4/5
Weight	4/5
Height	4/5
Age	3/5
Sex	2/5
Body Mass Index	2/5
Running Pace	2/5
Body Mass Index	2/5

a: Novice or non-novice runner, *b*: Recently Injured within 1-year: yes or no, *c*: Previously Injured: yes or no.

Table 9.12: The number of times each feature was selected for the 'Clinical feature trained Multi-Layer Perceptron model' as part of the hyperparameters search for the best 5 model iterations during Nested Cross Validation. The best 5 models were used to calculate the average accuracy of the Multi-Layer Perceptron model (0.58 AUC-ROC), and the following features appeared a number of times across the best models.

Feature	Feature Importance
Recently Injured[a]	5/5
Previously Injured[b]	5/5
Weight	5/5
Height	5/5
Running Distance	4/5
Running Experience[c]	4/5
Age	2/5
Body Mass Index	2/5
Running Pace	2/5
Sex	1/5

a: Recently Injured within 1-year: yes or no, *b*: Previously Injured: yes or no, *c*: Novice or non-novice runner.

Additional Results 1

As the Clinical feature dataset lead to the majority of higher AUC-ROC scores, a comparison is included in figures 9.10 and 9.11 (for approach 'a' and 'b' separately) with the three model training methods implemented in this thesis: (i) Default Model hyperparameters (Appendix section 11.5), (ii) Using the top 20 feature set with hyperparameter tuning (Appendix section 11.6) and (iii) Best feature set with hyperparameter tuning. These methods were implemented on all datasets and classifiers included in the Appendix. Training methods with Default Model hyperparameters only meant using only the default values for the hyperparameters, except for the case of the bagged SVM where a linear kernel model was fixed as the default in order to keep the model structure comparable to the other two training methods. Top 20 feature set with hyperparameter tuning meant using all of the top 20 features for each dataset as identified by SelectKbest, with the remaining hyperparameter values found via random search for the best combination. Using the best feature set with hyperparameter tuning meant letting both the number of top features used and all model hyperparameters be selected by the random search approach to find the best combination of both.

It can be seen for both approaches 'a' and 'b' generally for most classifiers, hyperparameter tuning lead to better overall accuracy, however results were classifier dependent where the other approaches may have performed better and additionally there were small differences in magnitude between the three methods. Full results for all classifiers and datasets for (i), (ii) and (iii) can be found in the appendix sections 11.5, 11.6 and 11.4, respectively.
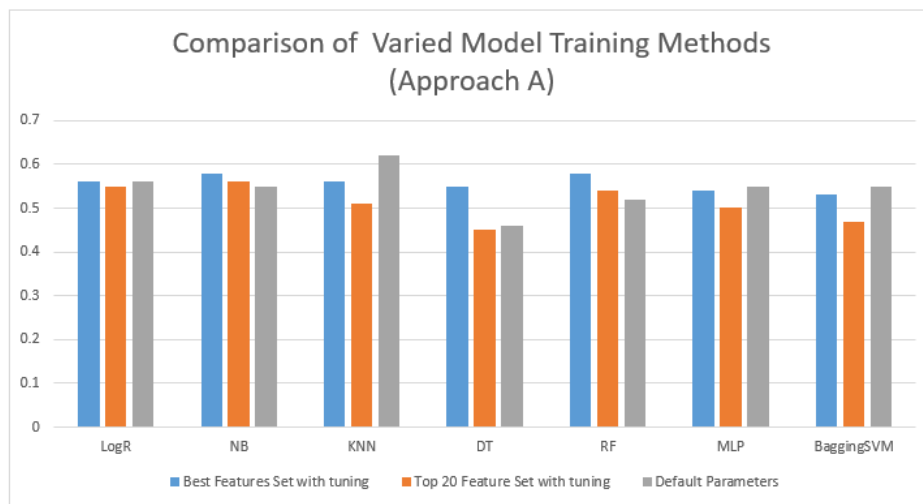
Figure 9.10: Comparison of the three model training methods carried out on the Clinical dataset for approach A.
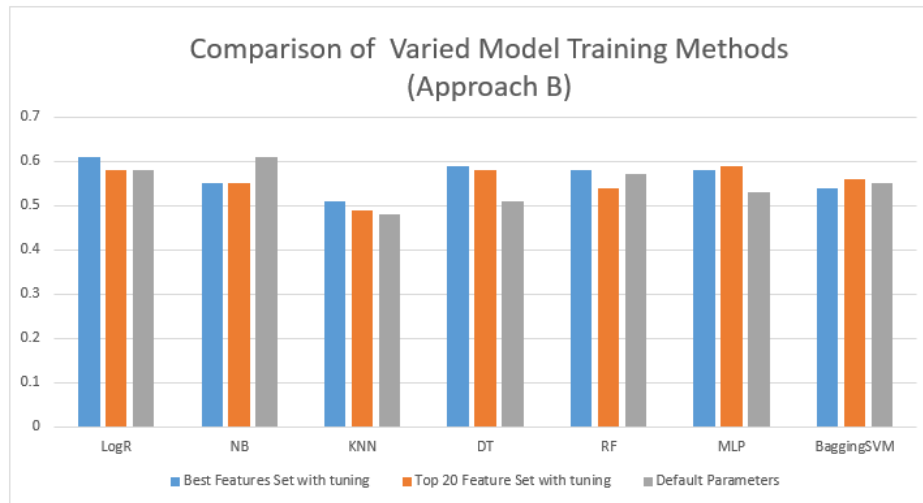
Figure 9.11: Comparison of the three model training methods carried out on the Clinical dataset for approach B.

Additional Results 2

Machine learning was used as a tool to examine the injury prediction problem in this chapter as a means to possibly discover a multivariate pattern linking our features with the injured and non-injured class data. Traditionally, analyses to investigate whether a relationship exists with-in features can be carried out following basic statistical methods. It is interesting that the implementation of Logistic regression was in the top two models for both approach 'a' and approach 'b'. A short statistical analysis was examined by performing a standard backwards step logistic regression implementation in SPSS on our Clinical dataset (best performing) with the best performing features identified in tables 9.7 and 9.8 using approach 'a'. This resulted in a 'Correct Classification Percentage' of 0.60%. Due to the fact all samples are used in the SPSS logistic regression analysis and there is no independent test set, this 0.60% is similar to training accuracy, which our logistic regression using sklearn scored 0.63% training accuracy on the same data and features. The SPSS results for Cox & Snell R Square and Nagelkerke R Square tests showed that the model only accounted for 10-13% of the explained variance.

9.2.4 Discussion

Despite an initial hypothesis that inertial data (impact accelerations or gyroscope) would be related to injury, the results indicated that the associated features resulted in poor models for predicting injury. Similarly, features from joint angle technique-based data also resulted in poor predictive

models. Almost all of the trained models achieved a poor prediction accuracy of <0.55 AUC-ROC. The most useful features and feature set proved to be the clinical information dataset (C). It can be seen that clinical features scored higher for both approaches: 'a': One sample per Subject, and 'b': One Sample per Stride. The weak predictive ability is reflected in the univariate statistical analyses where there were mainly very weak to weak effect sizes for the relationship between the features and injury outcome.

With regard to the aim to compare the use of the different datasets, specifically inertial features alone versus the use of a combination of inertial and kinematic features to predict injury, both the inertial feature only and the combined dataset with inertial and kinematic features performed poorly, creating weak predictive classifiers. Therefore, it seems (at least for the features used in this thesis) that using additional kinematic technique-based features did not add any additional benefit to the predictive models.

With regard to the second aim of determining which machine learning classifier is more effective at identifying those who are predisposed to injury, the average prediction score for each individual classifier tested on all datasets resulted in poor prediction accuracy (0.50 AUC) for all classifiers. As all classifiers performed poorly on average across datasets, it is difficult to reliably state which classifier performed best for this predictive problem. However, the clinical dataset performed the best out of the 6 datasets for individual models, achieving accuracies of 0.58 -0.61 AUC using Logistic Regression, Decision Tree and Multi-Layer Perceptron classifiers.

Feature importance was performed for the top 2 models in both approach 'a' and approach 'b', where it could be clearly seen that certain features were selected multiple times for model training. The top 6 important clinical features were identified as running experience (novice or not), previous injury status, running distance, weight and height. This is supported by previous studies that have found these factors to be related to injury, where both greater [34] and lower [132] levels of running experiences have been identified as possible risk factors for RRI's. With respect to previous injury, many studies have found a significantly greater risk of re-injury associated with having had a previous running injury [34, 132]. Running distance is a feature representative of training load which has some evidence of a risk factor for RRIs. It has been suggested that further training distance leads to a significantly greater risk for general RRI [35, 156]. With respect to weight and height, [270] and [274] found height to be an important variable for the prediction of lower limb injuries. Increased weight in women and decreased weight in men has also been identified as a possible a risk factor for foot and

back injuries [435].

In line with the weak predictive capabilities of the models within the current study, some previous studies have also found weak predictions [272,277]. [277] examined injury risk in youth male footballers using measures of neuromuscular tests with features such as asymmetry in the SLCMJ, 75% Hop and Y-balance, plus tuck jump knee valgus and anthropometrics. They found that their decision tree was a weak classifier with 0.66 AUC. [272] examined running injuries using kinematic measures collected during running and found that their random forest classifier had weak accuracy of 0.71%.

While three studies [275,276,278] found higher accuracies of prediction (approximately 0.75 AUC), they examined measures of strength and balance, rather than running technique. The reason for strength producing greater prediction perhaps is due to injuries being related to both loading and tissue integrity, with increased strength measures relating to increased tissue integrity. Similarly, [274] examined clinical and neuromuscular tests using features of strength and balance tests to achieve prediction accuracy in the range of 0.69-0.79 AUC.

One study has reported strong predictive capabilities [273], but used different measures from the current study. [273], trained a random forest classifier (0.8 AUC) on a number of kinematic parameters (Pelvic obliquity, Contralateral pelvic drop, etc), where only 2 of the top 20 features used were included in our model (Ankle dorsiflexion, Knee flexion). It is clear that the definition of the inertial features and kinematic features presented in this thesis may not have been the most effective. However, the small sample size used in Martinez et al. [273] (n=19) calls into question the accuracy and the generalisability of the model for other datasets/groups of runners. In addition, it was not clear in Martinez et al. at what stage the participant's injuries occurred, whether this was before or after the baseline data collection.

The area of prediction in running/sport injuries is challenging. The task of identifying factors which can be predictive of injury is far more difficult than identifying factors associated with injury [416]. Even in the area of ACL injuries associated with landing, which are one of the most widely examined, the ability to predict ACL injury is low, with Bahr suggesting that ''screening test ... do not work and probably never will''. This was in response to a study by [436] which suggested that neuromuscular control and valgus loading can accurately prospectively predict ACL injury in female athletes. However, while both [416] and [437] agree that there is a high association between injury and both neuromuscular control and valgus loading, they both conclude that these measures do not predict injury.

There are possibly three reasons to explain the low predictability found in the present study: (i) single acute measure, (ii) different injuries having different predictors, and (iii) different clusters of runners (with different running technique, irrespective of injury).

Firstly, it may not appropriate to take a single acute measure of a runner's loading and movement technique to be a representative example of their future running technique which may cause their future injury [438]. Similar assumptions are also made in Messier et al. [90] and Hein et al. [439]. A more appropriate approach would be to measure loading and running technique more frequently, ideally on a run-by-run basis.

Secondly, subjects prospectively sustained a number of different running related injuries (e.g. Plantar Fasciitis, Calf Strain, Achilles Tendinopathy etc.) but for the purpose of this thesis, all data were combined together for the binary classification of injury versus no injury. It is possible that different injuries are caused by different loading and technique features, which when combined together mask each other. For example, Hip flexion ROM may be related to hamstring injury [262], and knee abduction moment may be related to ACL injury [436]. This suggestion is supported by the number of studies which reported higher levels of predictive accuracy when identifying the risk of ACL injury (95- 97% [265, 266, 269]) compared to studies which focused on identifying the risk of general running related injuries (0.69-0.75 AUC [272, 276, 277]). Although it may have been more difficult to predict general injury versus a specific injury (e.g. Plantar Fasciitis), the decision was made to attempt to predict general injuries in the present study as a reliably accurate prediction of general injury could be extremely beneficial to all runners.

Thirdly, different runners have different running techniques irrespective of injury type. Therefore, it is possible that different running techniques have different causative factors for injury and would require different predictive models. With respect to athletes with the same injury, Franklyn et al. [440] found when performing agility turns to evaluate groin-based injuries, that athletes could be separated into three distinct clusters which represented three different movement strategies adopted by athletes with groin pain. Similarly, with respect to running gait patterns in healthy pain-free subjects, Phinyomark et al. [441] was able to identify two distinct running gait patterns.

The aim of this research was to be able to build a machine learning model which would be able to predict whether a runner was likely to become injured based on their loading and movement technique during running. The implications of not being able to build an accurate or reliable model are that we cannot identify runners who may become injured, and we cannot focus solely on specific runners

to target for injury prevention strategies. Without an approach that is capable of identifying those runners who are more at risk of a RRI, we must continue to target all runners and encourage the uptake of injury prevention interventions or strategies [416].

9.2.5 Limitations

One limitation of our study was the decision to group all injury types together in the dataset in order to predict a general running related injury. It is possible that binary classification of specific injury (e.g. Calf Strain) versus no injury, or multi-class classification of specific injury 1 (e.g. Calf Strain) versus specific injury 2 (e.g. Plantar Fasciitis) versus no injury, may result in better predictive capabilities due to the uncoupling of the injuries and the features which may be able to distinguish specific injuries [262, 269]. A further limitation is in the experimental set-up, in which data were only collected as a once off acute measurement and this was deemed to be representative of their running loading and technique for the subsequent 12-months. Alenezi et al. [438] found that kinematic variables can suffer from greater variability when measured on multiple days compared to measures of vGRFs. This could be improved by collecting running data on multiple occasions at regular intervals over the course of the observation period, and so if a runner became injured, their most recently collected data would be more representative of the level of loading and technique during their running style which may have contributed to injury. A final limitation was in how all the data collected for model building was captured on a treadmill in a laboratory environment, which may not be reflective of overground running where most runners train and incur injuries [442].

9.2.6 Future Work

Future work could examine a number of areas. Firstly, it could investigate how identification of injury may be improved by taking multiple measures of a person's loading and running technique over time. Secondly, a potentially successful approach for future work may be to split the original dataset into specific injury types (e.g. Plantar Fasciitis, Calf Strain, Achilles Tendinopathy etc.,) and to use machine learning to predict likelihood of specific injury rather than general RRI as examined out in this chapter. There may be variables/features that prove to have strong predictive capability for specific injuries that may have been masked by considering all injury types together. Finally, the features used in this thesis were discrete in nature; potentially more useful features could be found by examining and extracting continuous features of the time-series data or using functional data analysis

approaches [272].

9.2.7   Conclusion

Study 3 explored whether an accurate injury prediction model could be built which would be able to predict whether runners were likely to become injured or not. The results showed that the machine learning models built using a range of approaches and different input features could not confidently predict the likelihood of injury. This does not mean that these features/variables are not related to injury, but in the form used in this thesis they were not predictive of injury. This may suggest that it is not appropriate to take a single acute measure of a runner's loading and movement technique to be a representative example of future injury status. In addition, it was shown the inclusion of kinematic technique-based features appeared to be of no additional benefit to the inertial features for injury prediction, both datasets had poor predictive ability. The clinical dataset produced the most accurate classifiers, where important clinical features were identified by the models. Running experience, previous injury status, running distance, weight and height were identified as being important features to consider. As we were unable to predict runners likely to become injured, and because so many people can become injured (49% of our 150 runners in the experiment became injured), it is advised that all runners are encouraged to take appropriate precautions to prevent RRIs [416], until there is a way of providing more focused feedback on injury predisposition. Such approaches have been taken to prevent injuries in other sports (e.g. FIFA 11 soccer [443], GAA 15 in Gaelic sports [444]).

Chapter 10

Conclusion

10.1 Overall Discussion and Conclusion

The aim of this body of research was overall to better examine technique and loading in running for the purpose of understanding and preventing running related injuries. This research addressed a number of areas within this broader aim, namely: (i) to understand and improve the level of accuracy of 6DOF estimation of running technique (Study 1), (ii) to compare the effectiveness of loading-based and technique-based measures of biofeedback via a mobile-based system to reduce loading on the body, to reduce risk of RRI (Study 2), and (iii) to develop accurate models to predict individuals who are at greater risk of RRIs and use these models for the identification of RRI causative factors (Study 3).

In Study 1.1 (Section 7.3, Chapter 7), first the baseline accuracy was established by implementing Madgwick's 6DOF orientation filter [14] to understand the level of orientation error that existed during running. The level of error in the orientation estimates were far too high to be used for clinical assessments, with RMSE greater than 10 degrees on most segment axes. In comparison to previous studies (with errors <6°) [8, 211], it can be seen that this level of error is too high for accurate analysis of running technique. This was the first study to quantify the level of inaccuracies present in 6DOF running orientation over a duration similar to a training session (30 minutes) and was the first study to evaluate removing Z-axis drift via linear de-drifting. The implication of this was that additional approaches were needed to be investigated to reduce the orientation error to acceptable levels. These errors were not solely due to signal drift and 6DOF estimation as they were not consistent and were seemingly random. Further investigation discovered the presence of data loss which caused the orientation filter to be disturbed and affect convergence of filter estimates. To address this, two main approaches were explored with the aim of improving the convergence of the algorithm and in doing so reducing the orientation error. These approaches involved (i) investigating the use of a pre-processing pipeline to pre-fill and post-fill the missing data (Study 1.2) and (ii) additionally the exploration of

two Madgwick filter modifications (Study 1.3A and Study 1.3B). With respect to data-loss, pre-filling the missing inertial data using linear interpolation or previous sample filling was found not to have a significant impact on improving the orientation errors. This is the first study to demonstrate the disruptive impact of data loss on 6DOF orientation estimation and to investigate the effectiveness of pre-filling and post-filling missing data across all sensor axes. The implication of the failure of this pre-processing pipeline to reduce orientation errors is that missing inertial data will still pose problems for 6DOF estimation. It is possible that more sophisticated gap-filling approaches (e.g., machine learning-based solutions) may be beneficial.

With regards to the two Madgwick filter modifications, the first filter modification which altered the beta parameter, has been shown in other studies to improve orientation estimates for both 9DOF [12, 366] and 6DOF estimation [7, 216]. An optimised beta parameter of 0.1 was found to improve the orientation error across all segments (tibia, thigh and sacrum) in the present study. Despite beta parameter tuning leading to more accurate results compared to the baseline (average reduction in error of 4-6°, 2° and 1-2° across the X, Y, Z axes of the data, respectively), the orientation errors were still too high to be used for absolute orientation estimation in a clinical setting. The second modification, which was the introduction of multiple iterations of Stochastic Gradient Descent and the inclusion of a momentum parameter, has not been previously examined in the literature. A grid search was employed to find the best combination of parameters in order to reduce the orientation error. It was found that the best set of parameters was different for each segment, and for each beta parameter tested (baseline = 0.033 and optimised = 0.1). Generally, the best parameter combinations resulted in slightly reduced orientation error, however this reduction in orientation error was in generally not significant and had mixed results for the optimised beta set BestParam0.1. This suggests that out of the two modified approaches, the simple beta parameter tuning (which requires less computation) is the recommended approach for future parameter tuning.

Although the improvements in accuracy were not sufficient enough to allow clinical measurements of an absolute nature; it was found that this improvement was sufficiently accurate to successfully provide information about relative changes in technique with respect to a person's own baseline values. This finding was utilised in Chapter 8 (Study 2) where both impact loading-based (impact acceleration) and technique-based (thigh angle) biofeedback were provided to two groups of recreational runners in an attempt to reduce loading on the body at the tibia and at the sacrum. It was found that both forms of biofeedback were effective at reducing impact loading on the body, however, loading-based feedback

(impact accelerations) had a greater reduction of loading at the tibia compared to the sacrum, and technique-based feedback (thigh angle) had a greater reduction at the sacrum compared to the tibia. As both forms of biofeedback were effective at reducing impact loading on the body, it may be advisable for future studies to provide tibia-based tibial acceleration biofeedback as it is easier to secure a sensor to this location compared to the thigh. This was the first study to directly compare loading-based and technique-based biofeedback as measured via inertial sensors. In addition, this appears to be the first study to investigate the provision of inertial sensor-based estimation of knee kinematics for biofeedback during running. While it is not possible to conclusively state whether the location of where the biofeedback was provided (the tibia versus the thigh) or the form of biofeedback itself (impact accelerations versus thigh angle) had a greater influence on the greater reduction in impact loading on the body, it is possible that the differences may be due to how the different forms of biofeedback lead to different changes in a runner's technique. The tibia-based feedback may have resulted in more in ankle dorsi-flexion, and less centre of mass and foot vertical velocity at initial contact [250], whereas thigh-based feedback may have resulted in more knee and hip flexion at initial contact [246], since direct feedback of knee kinematics were provided to the runner. This study provides a very valuable insight to running retraining for injury prevention, and the system app-based biofeedback system developed for this thesis provides a relatively cheap, fully portable and novel technology solution for runners.

Finally in Chapter 9 (Study 3), a set of experiments aimed to build a machine learning running injury prediction model using prospective injury data. Overall, the majority of trained models had poor predictive accuracy <0.55 AUC. This was the first study to analyse directly whether kinematic data features provided any additional predictive benefit to using inertial data features only for the purpose of predicting running injuries. It was found that the kinematic technique-based features did not provide additional prediction benefits to the inertial-based feature set used in the present study. On average, classifiers performed similarly poorly when applied to all six datasets, and so on average there was not a clear single best performing classifier. However, on an individual model basis, certain implementations of Logistic Regression, Decision Tree and Multi-Layer Perceptron classifiers with the clinical dataset performed better. These were weak predictive models with top model scores of 0.58-0.61 AUC. Weak predictive models for the classification of general running injuries have also been found in other studies [272, 277]. However other studies [273] have found high levels of injury prediction (0.8 AUC) by primarily using features different from the present study (Pelvic obliquity,

Contralateral pelvic drop, etc.). However the generalizability of the findings in Martinez et al. [273] are questionable due to the small sample size. It is possible that the features used in this study were not sufficiently detailed to classify general running injury and that the mixture of data from different specific injuries in the dataset may have further masked their usefulness. Overall, the performance of the clinical dataset outperformed the other five feature datasets. Feature importance identified a number of clinical features (running experience, previous injury, running distance, weight and height) which were deemed important by the classifiers. This is supported by previous research which has linked each of these factors with the risk of RRI's (running experience [34], previous injury [132], running distance [156], weight [435] and height [270]). The implications of not being able to build a reliable prediction model are that we cannot identify runners who may become injured, and therefore we must continue to target all runners and encourage the uptake of injury prevention interventions [416]. Future research in each of the areas of study are provided in the relevant section (Section 7.6.5, Section 8.4.4 and Section 9.2.6).

Bibliography

[1] H. van der Worp, J. W. Vrielink, and S. W. Bredeweg, "Do runners who suffer injuries have higher vertical ground reaction forces than those who remain injury-free? a systematic review and meta-analysis," *British journal of sports medicine*, vol. 50, no. 8, pp. 450–457, 2016.

[2] S. A. Dugan and K. P. Bhat, "Biomechanics and analysis of running gait," *Physical Medicine and Rehabilitation Clinics*, vol. 16, no. 3, pp. 603–621, 2005.

[3] P. Date, *Combinatorial neural network training algorithm for neuromorphic computing.* Rensselaer Polytechnic Institute, 2019.

[4] D.-c. Lee, A. G. Brellenthin, P. D. Thompson, X. Sui, I.-M. Lee, and C. J. Lavie, "Running as a key lifestyle medicine for longevity," *Progress in cardiovascular diseases*, vol. 60, no. 1, pp. 45–55, 2017.

[5] R. Van Gent, D. Siem, M. van Middelkoop, A. Van Os, S. Bierma-Zeinstra, and B. Koes, "Incidence and determinants of lower extremity running injuries in long distance runners: a systematic review," *British journal of sports medicine*, vol. 41, no. 8, pp. 469–480, 2007.

[6] A. M. Sabatini, "Estimating three-dimensional orientation of human body parts by inertial/magnetic sensing," *Sensors*, vol. 11, no. 2, pp. 1489–1525, 2011.

[7] H. Chen, M. C. Schall Jr, and N. Fethke, "Accuracy of angular displacements and velocities from inertial-based inclinometers," *Applied ergonomics*, vol. 67, pp. 151–161, 2018.

[8] P. B. Shull, J. Xu, B. Yu, and X. Zhu, "Magneto-gyro wearable sensor algorithm for trunk sway estimation during walking and running gait," *IEEE Sensors Journal*, vol. 17, no. 2, pp. 480–486, 2016.

[9] M. El-Gohary and J. McNames, "Human joint angle estimation with inertial sensors and validation with a robot arm," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 7, pp. 1759–1767, 2015.

[10] Q. Yuan, E. Asadi, Q. Lu, G. Yang, and I.-M. Chen, "Uncertainty-based imu orientation tracking algorithm for dynamic motions," *IEEE/ASME Transactions on Mechatronics*, vol. 24, no. 2, pp. 872–882, 2019.

[11] C. Yi, J. Ma, H. Guo, J. Han, H. Gao, F. Jiang, and C. Yang, "Estimating three-dimensional body orientation based on an improved complementary filter for human motion tracking," *Sensors*, vol. 18, no. 11, p. 3765, 2018.

[12] E. Bergamini, G. Ligorio, A. Summa, G. Vannozzi, A. Cappozzo, and A. M. Sabatini, "Estimating orientation using magnetic and inertial sensors and different sensor fusion approaches: Accuracy assessment in manual and locomotion tasks," *Sensors*, vol. 14, no. 10, pp. 18625–18649, 2014.

[13] A. D. Young, M. J. Ling, and D. Arvind, "Orient-2: A realtime wireless posture tracking system using local orientation estimation," in *Proceedings of the 4th workshop on Embedded networked sensors*, pp. 53–57, 2007.

[14] S. O. Madgwick, A. J. Harrison, and R. Vaidyanathan, "Estimation of imu and marg orientation using a gradient descent algorithm," in *2011 IEEE international conference on rehabilitation robotics*, pp. 1–7, IEEE, 2011.

[15] E. Sihite and T. Bewley, "Attitude estimation of a high-yaw-rate mobile inverted pendulum; comparison of extended kalman filtering, complementary filtering, and motion capture," in *2018 Annual American Control Conference (ACC)*, pp. 5831–5836, IEEE, 2018.

[16] S. Shan, Z. Hou, and J. Wu, "Linear kalman filter for attitude estimation from angular rate and a single vector measurement," *Journal of Sensors*, vol. 2017, 2017.

[17] S. Yang, C. Mohr, and Q. Li, "Ambulatory running speed estimation using an inertial sensor," *Gait & posture*, vol. 34, no. 4, pp. 462–466, 2011.

[18] J. B. Lee, K. J. Sutter, C. D. Askew, and B. J. Burkett, "Identifying symmetry in running gait using a single inertial sensor," *Journal of Science and Medicine in Sport*, vol. 13, no. 5, pp. 559–563, 2010.

[19] M. Caruso, A. M. Sabatini, M. Knaflitz, M. Gazzoni, U. Della Croce, and A. Cereatti, "Orientation estimation through magneto-inertial sensor fusion: A heuristic approach for suboptimal parameters tuning," *IEEE Sensors Journal*, vol. 21, no. 3, pp. 3408–3419, 2020.

[20] Z. Y. Chan, J. H. Zhang, I. P. Au, W. W. An, G. L. Shum, G. Y. Ng, and R. T. Cheung, "Gait retraining for the reduction of injury occurrence in novice distance runners: 1-year follow-up of a randomized controlled trial," *The American journal of sports medicine*, vol. 46, no. 2, pp. 388–395, 2018.

[21] B. J. Bowser, R. Fellin, C. E. Milner, M. B. Pohl, and I. S. Davis, "Reducing impact loading in runners: A one-year follow-up," *Medicine and science in sports and exercise*, vol. 50, no. 12, p. 2500, 2018.

[22] K. R. Williams and P. R. Cavanagh, "Relationship between distance running mechanics, running economy, and performance," *Journal of Applied Physiology*, vol. 63, no. 3, pp. 1236–1245, 1987.

[23] H. P. Crowell and I. S. Davis, "Gait retraining to reduce lower extremity loading in runners," *Clinical biomechanics*, vol. 26, no. 1, pp. 78–83, 2011.

[24] M. Baggaley, R. Willy, and S. Meardon, "Primary and secondary effects of real-time feedback to reduce vertical loading rate during running," *Scandinavian journal of medicine & science in sports*, vol. 27, no. 5, pp. 501–507, 2017.

[25] L. M. van Gelder, A. Barnes, J. S. Wheat, and B. W. Heller, "The use of biofeedback for gait retraining: A mapping review," *Clinical Biomechanics*, vol. 59, pp. 159–166, 2018.

[26] C. Ó Catháin, C. Richter, and K. Moran, "Can directed compliant running reduce the magnitude variables associated with the development of running injuries?," *Journal of strength and conditioning research*, 2020.

[27] C. E. Milner, R. Ferber, C. D. Pollard, J. Hamill, and I. S. Davis, "Biomechanical factors associated with tibial stress fracture in female runners," *Medicine and science in sports and exercise*, vol. 38, no. 2, p. 323, 2006.

[28] M. J. Duffey, D. F. Martin, D. W. Cannon, T. Craven, and S. P. Messier, "Etiologic factors associated with anterior knee pain in distance runners.," *Medicine and science in sports and exercise*, vol. 32, no. 11, pp. 1825–1832, 2000.

[29] D. L. Goss and M. T. Gross, "Relationships among self-reported shoe type, footstrike pattern, and injury incidence.," *US Army Medical Department Journal*, 2012.

[30] K. Hollander, C. D. Johnson, J. Outerleys, and I. S. Davis, "Multifactorial determinants of running injury locations in 550 injured recreational runners.," *Medicine and Science in Sports and Exercise*, 2020.

[31] I. S. Davis, B. J. Bowser, and D. R. Mullineaux, "Greater vertical impact loading in female runners with medically diagnosed injuries: a prospective investigation," *British journal of sports medicine*, vol. 50, no. 14, pp. 887–892, 2016.

[32] R. Ferber, I. M. Davis, J. Hamill, C. D. Pollard, and K. A. McKeown, "Kinetic variables in subjects with previous lower extremity stress fractures," *Medicine & Science in Sports & Exercise*, vol. 34, no. 5, p. S5, 2002.

[33] C. Bramah, S. J. Preece, N. Gill, and L. Herrington, "Is there a pathological gait associated with common soft tissue running injuries?," *The American journal of sports medicine*, vol. 46, no. 12, pp. 3023–3031, 2018.

[34] D. Y. Wen, J. C. Puffer, and T. P. Schmalzried, "Injuries in runners: a prospective study of alignment.," *Clinical journal of sport medicine: official journal of the Canadian Academy of Sport Medicine*, vol. 8, no. 3, pp. 187–194, 1998.

[35] K. Knobloch, U. Yoon, and P. M. Vogt, "Acute and overuse injuries correlated to hours of training in master running athletes," *Foot & ankle international*, vol. 29, no. 7, pp. 671–676, 2008.

[36] J. B. Morris, D. L. Goss, E. M. Miller, and I. S. Davis, "Using real-time biofeedback to alter running biomechanics: a randomized controlled trial," *Translational Sports Medicine*, vol. 3, no. 1, pp. 63–71, 2020.

[37] J. P. Holden and P. R. Cavanagh, "The free moment of ground reaction in distance running and its changes with pronation," *Journal of biomechanics*, vol. 24, no. 10, pp. 887–897, 1991.

[38] J. T. Kalkhoven, M. L. Watsford, and F. M. Impellizzeri, "A conceptual model and detailed framework for stress-related, strain-related, and overuse athletic injury," *Journal of science and medicine in sport*, vol. 23, no. 8, pp. 726–734, 2020.

[39] K. R. Williams, "Biomechanics of running," *Exercise and sport sciences reviews*, vol. 13, no. 1, pp. 389–442, 1985.

[40] C. F. Munro, D. I. Miller, and A. J. Fuglevand, "Ground reaction forces in running: a reexamination," *Journal of biomechanics*, vol. 20, no. 2, pp. 147–155, 1987.

[41] A. Hreljac, "Impact and overuse injuries in runners," *Medicine and science in sports and exercise*, vol. 36, no. 5, pp. 845–849, 2004.

[42] T. Ueda, H. Hobara, Y. Kobayashi, T. Heldoorn, M. Mochimaru, and H. Mizoguchi, "Comparison of 3 methods for computing loading rate during running," *International Journal of Sports Medicine*, vol. 37, no. 13, pp. 1087–1090, 2016.

[43] R. T. Cheung and M. J. Rainbow, "Landing pattern and vertical loading rates during first attempt of barefoot running in habitual shod runners," *Human movement science*, vol. 34, pp. 120–127, 2014.

[44] A. A. Zadpoor and A. A. Nikooyan, "The relationship between lower-extremity stress fractures and the ground reaction force: a systematic review," *Clinical biomechanics*, vol. 26, no. 1, pp. 23–28, 2011.

[45] A. Tongen and R. E. Wunderlich, "Biomechanics of running and walking," *Mathematics and Sports*, vol. 43, pp. 1–12, 2010.

[46] T. F. Novacheck, "The biomechanics of running," *Gait & posture*, vol. 7, no. 1, pp. 77–95, 1998.

[47] A. G. Schache, K. L. Bennell, P. D. Blanch, and T. V. Wrigley, "The coordinated movement of the lumbo–pelvic–hip complex during running: a literature review," *Gait & posture*, vol. 10, no. 1, pp. 30–47, 1999.

[48] M. Pink, J. Perry, P. A. Houglum, and D. J. Devine, "Lower extremity range of motion in the recreational sport runner," *The American journal of sports medicine*, vol. 22, no. 4, pp. 541–549, 1994.

[49] F. L. Buczek and P. R. Cavanagh, "Stance phase knee and ankle kinematics and kinetics during level and downhill running.," *Medicine and science in sports and exercise*, vol. 22, no. 5, pp. 669–677, 1990.

[50] F. Tian, N. Li, Z. Zheng, Q. Huang, T. Zhu, Q. Li, W. Wang, T.-Y. Tsai, and S. Wang, "The effects of marathon running on three-dimensional knee kinematics during walking and running in recreational runners," *Gait & Posture*, vol. 75, pp. 72–77, 2020.

[51] H.-L. Teng and C. M. Powers, "Influence of trunk posture on lower extremity energetics during running," *Med Sci Sports Exerc*, vol. 47, no. 3, pp. 625–630, 2015.

[52] D. A. Winter and P. J. Bishop, "Lower extremity injury," *Sports Medicine*, vol. 14, no. 3, pp. 149–156, 1992.

[53] W. B. Edwards, "Modeling overuse injuries in sport as a mechanical fatigue phenomenon," *Exercise and sport sciences reviews*, vol. 46, no. 4, pp. 224–231, 2018.

[54] A. Hreljac, "Etiology, prevention, and early intervention of overuse injuries in runners: a biomechanical perspective," *Physical Medicine and Rehabilitation Clinics*, vol. 16, no. 3, pp. 651–667, 2005.

[55] B. Ewers, D. Dvoracek-Driksna, M. Orth, and R. Haut, "The extent of matrix damage and chondrocyte death in mechanically traumatized articular cartilage explants depends on rate of loading," *Journal of Orthopaedic Research*, vol. 19, no. 5, pp. 779–784, 2001.

[56] M. Bertelsen, A. Hulme, J. Petersen, R. K. Brund, H. Sørensen, C. Finch, E. T. Parner, and R. Nielsen, "A framework for the etiology of running-related injuries," *Scandinavian journal of medicine & science in sports*, vol. 27, no. 11, pp. 1170–1180, 2017.

[57] T. Soligard, M. Schwellnus, J.-M. Alonso, R. Bahr, B. Clarsen, H. P. Dijkstra, T. Gabbett, M. Gleeson, M. Hägglund, M. R. Hutchinson, *et al.*, "How much is too much?(part 1) international olympic committee consensus statement on load in sport and risk of injury," *British journal of sports medicine*, vol. 50, no. 17, pp. 1030–1041, 2016.

[58] M. P. Van der Worp, D. S. Ten Haaf, R. van Cingel, A. de Wijer, M. W. Nijhuis-van der Sanden, and J. B. Staal, "Injuries in runners; a systematic review on risk factors and sex differences," *PloS one*, vol. 10, no. 2, p. e0114937, 2015.

[59] P. W. Kline and D. B. Williams III, "Effects of normal aging on lower extremity loading and coordination during running in males and females," *International journal of sports physical therapy*, vol. 10, no. 6, p. 901, 2015.

[60] R. Meeusen, M. Duclos, M. Gleeson, G. Rietjens, J. Steinacker, and A. Urhausen, "Prevention, diagnosis and treatment of the overtraining syndrome," *European Journal of Sport Science*, vol. 6, no. 1, pp. 1–14, 2006.

[61] S. K. Grimston, J. R. Engsberg, R. Kloiber, and D. A. Hanley, "Bone mass, external loads, and stress fracture in female runners," *Journal of Applied Biomechanics*, vol. 7, no. 3, pp. 293–302, 1991.

[62] A. Hreljac, R. N. Marshall, and P. A. Hume, "Evaluation of lower extremity overuse injury potential in runners," *Medicine & Science in Sports & Exercise*, vol. 32, no. 9, pp. 1635–1641, 2000.

[63] T. P. Franke, F. J. Backx, and B. M. Huisstede, "Running themselves into the ground? incidence, prevalence, and impact of injury and illness in runners preparing for a half or full marathon," *journal of orthopaedic & sports physical therapy*, vol. 49, no. 7, pp. 518–528, 2019.

[64] T. P. Yamato, B. T. Saragiotto, and A. D. Lopes, "A consensus definition of running-related injury in recreational runners: a modified delphi approach," *Journal of orthopaedic & sports physical therapy*, vol. 45, no. 5, pp. 375–380, 2015.

[65] S. W. Bredeweg, B. Kluitenberg, B. Bessem, and I. Buist, "Differences in kinetic variables between injured and noninjured novice runners: a prospective cohort study," *Journal of Science and Medicine in Sport*, vol. 16, no. 3, pp. 205–210, 2013.

[66] A. Bovens, G. Janssen, H. Vermeer, J. Hoeberigs, M. Janssen, and F. Verstappen, "Occurrence of running injuries in adults following a supervised training program," *International Journal of Sports Medicine*, vol. 10, no. S 3, pp. S186–S190, 1989.

[67] S. N. Blair, H. W. Kohl, and N. N. Goodyear, "Rates and risks for running and exercise injuries: studies in three populations," *Research Quarterly for Exercise and Sport*, vol. 58, no. 3, pp. 221–228, 1987.

[68] R. I. Dudley, D. N. Pamukoff, S. K. Lynn, R. D. Kersey, and G. J. Noffal, "A prospective comparison of lower extremity kinematics and kinetics between injured and non-injured collegiate cross country runners," *Human movement science*, vol. 52, pp. 197–202, 2017.

[69] C. Napier, C. MacLean, J. Maurer, J. Taunton, and M. Hunt, "Kinetic risk factors of running-related injuries in female recreational runners," *Scandinavian journal of medicine & science in sports*, vol. 28, no. 10, pp. 2164–2172, 2018.

[70] K. E. Gerlach, S. C. White, H. W. Burton, J. M. Dorn, J. J. Leddy, and P. J. Horvath, "Kinetic changes with fatigue and relationship to injury in female runners.," *Medicine and science in sports and exercise*, vol. 37, no. 4, pp. 657–663, 2005.

[71] D. J. Kuhman, M. R. Paquette, S. A. Peel, and D. A. Melcher, "Comparison of ankle kinematics and ground reaction forces between prospectively injured and uninjured collegiate cross country runners," *Human movement science*, vol. 47, pp. 9–15, 2016.

[72] J.-F. Esculier, J.-S. Roy, and L. J. Bouyer, "Lower limb control and strength in runners with and without patellofemoral pain syndrome," *Gait & posture*, vol. 41, no. 3, pp. 813–819, 2015.

[73] M. B. Pohl, J. Hamill, and I. S. Davis, "Biomechanical and anatomic factors associated with a history of plantar fasciitis in female runners," *Clinical Journal of Sport Medicine*, vol. 19, no. 5, pp. 372–376, 2009.

[74] J. L. McCRORY, D. F. Martin, R. B. Lowery, D. W. Cannon, W. W. Curl, H. M. Read Jr, D. M. Hunter, T. Craven, and S. P. Messier, "Etiologic factors associated with achilles tendinitis in runners.," *Medicine and science in sports and exercise*, vol. 31, no. 10, pp. 1374–1381, 1999.

[75] S. P. Messier, S. E. Davis, W. W. Curl, R. B. Lowery, and R. J. Pack, "Etiologic factors associated with patellofemoral pain in runners.," *Medicine and science in sports and exercise*, vol. 23, no. 9, pp. 1008–1015, 1991.

[76] L. B. Azevedo, M. I. Lambert, C. L. Vaughan, C. M. O'Connor, and M. P. Schwellnus, "Biomechanical variables associated with achilles tendinopathy in runners," *British journal of sports medicine*, vol. 43, no. 4, pp. 288–292, 2009.

[77] J. Bigouette, J. Simon, K. Liu, and C. L. Docherty, "Altered vertical ground reaction forces in participants with chronic ankle instability while running," *Journal of athletic training*, vol. 51, no. 9, pp. 682–687, 2016.

[78] A. P. Ribeiro, S. M. A. João, R. C. Dinato, V. D. Tessutti, and I. C. N. Sacco, "Dynamic patterns of forces and loading rate in runners with unilateral plantar fasciitis: A cross-sectional study," *PloS one*, vol. 10, no. 9, p. e0136971, 2015.

[79] K. R. Sheerin, D. Reid, and T. F. Besier, "The measurement of tibial acceleration in runners—a review of the factors that can affect tibial acceleration during running and evidence-based guidelines for its use," *Gait & Posture*, vol. 67, pp. 12–24, 2019.

[80] R. A. Zifchock, I. Davis, and J. Hamill, "Kinetic asymmetry in female runners with and without retrospective tibial stress fractures," *Journal of biomechanics*, vol. 39, no. 15, pp. 2792–2797, 2006.

[81] R. A. Zifchock, I. Davis, J. Higginson, S. McCaw, and T. Royer, "Side-to-side differences in overuse running injury susceptibility: a retrospective study," *Human movement science*, vol. 27, no. 6, pp. 888–902, 2008.

[82] A. I. Daoud, G. J. Geissler, F. Wang, J. Saretsky, Y. A. Daoud, and D. E. Lieberman, "Foot strike and injury rates in endurance runners: a retrospective study," *Med Sci Sports Exerc*, vol. 44, no. 7, pp. 1325–1334, 2012.

[83] M. Paquette, C. Milner, and D. Melcher, "Foot contact angle variability during a prolonged run with relation to injury history and habitual foot strike pattern," *Scandinavian journal of medicine & science in sports*, vol. 27, no. 2, pp. 217–222, 2017.

[84] J. Hamill and A. H. Gruber, "Is changing footstrike pattern beneficial to runners?," *Journal of sport and health science*, vol. 6, no. 2, pp. 146–153, 2017.

[85] M. O. de Almeida, B. T. Saragiotto, T. P. Yamato, and A. D. Lopes, "Is the rearfoot pattern the most frequently foot strike pattern among recreational shod distance runners?," *Physical Therapy in Sport*, vol. 16, no. 1, pp. 29–33, 2015.

[86] J.-P. Kulmala, J. Avela, K. Pasanen, and J. Parkkari, "Forefoot strikers exhibit lower running-induced knee loading than rearfoot strikers," *Med Sci Sports Exerc*, vol. 45, no. 12, pp. 2306–2313, 2013.

[87] A. R. Altman and I. S. Davis, "A kinematic method for footstrike pattern detection in barefoot and shod runners," *Gait & posture*, vol. 35, no. 2, pp. 298–300, 2012.

[88] D. Sugimoto, B. D. Kelly, D. L. Mandel, D. A. d'Hemecourt, S. C. Carpenito, C. A. d'Hemecourt, and P. A. d'Hemecourt, "Running propensities of athletes with hamstring injuries," *Sports*, vol. 7, no. 9, p. 210, 2019.

[89] B. J. Warr, R. E. Fellin, S. G. Sauer, D. L. Goss, P. N. Frykman, and J. F. Seay, "Characterization of foot-strike patterns: lack of an association with injuries or performance in soldiers," *Military medicine*, vol. 180, no. 7, pp. 830–834, 2015.

[90] S. P. Messier, D. F. Martin, S. L. Mihalko, E. Ip, P. DeVita, D. W. Cannon, M. Love, D. Beringer, S. Saldana, R. E. Fellin, *et al.*, "A 2-year prospective cohort study of overuse running injuries: the runners and injury longitudinal study (trails)," *The American journal of sports medicine*, vol. 46, no. 9, pp. 2211–2221, 2018.

[91] B. Dingenen, P. Malliaras, T. Janssen, L. Ceyssens, R. Vanelderen, and C. J. Barton, "Two-dimensional video analysis can discriminate differences in running kinematics between recreational runners with and without running-related knee injury," *Physical Therapy in Sport*, vol. 38, pp. 184–191, 2019.

[92] R. Mann, L. Malisoux, C. Nührenbörger, A. Urhausen, K. Meijer, and D. Theisen, "Association of previous injury and speed with running style and stride-to-stride fluctuations," *Scandinavian journal of medicine & science in sports*, vol. 25, no. 6, pp. e638–e645, 2015.

[93] O. A. Donoghue, A. J. Harrison, P. Laxton, and R. K. Jones, "Lower limb kinematics of subjects with chronic achilles tendon injury during running," *Research in Sports Medicine*, vol. 16, no. 1, pp. 23–38, 2008.

[94] T. R. Derrick, "The effects of knee contact angle on impact forces and accelerations.," *Medicine and science in sports and exercise*, vol. 36, no. 5, pp. 832–837, 2004.

[95] C. M. Powers, "The influence of altered lower-extremity kinematics on patellofemoral joint dysfunction: a theoretical perspective," *Journal of Orthopaedic & Sports Physical Therapy*, vol. 33, no. 11, pp. 639–646, 2003.

[96] M. L. Ireland, J. D. Willson, B. T. Ballantyne, and I. M. Davis, "Hip strength in females with and without patellofemoral pain," *Journal of orthopaedic & sports physical therapy*, vol. 33, no. 11, pp. 671–676, 2003.

[97] T. Q. Lee, G. Morris, and R. P. Csintalan, "The influence of tibial and femoral rotation on patellofemoral contact area and pressure," *Journal of Orthopaedic & Sports Physical Therapy*, vol. 33, no. 11, pp. 686–693, 2003.

[98] B. Noehren, I. Davis, and J. Hamill, "Asb clinical biomechanics award winner 2006: Prospective study of the biomechanical factors associated with iliotibial band syndrome," *Clinical biomechanics*, vol. 22, no. 9, pp. 951–956, 2007.

[99] C. E. Milner, J. Hamill, and I. Davis, "Are knee mechanics during early stance related to tibial stress fracture in runners?," *Clinical Biomechanics*, vol. 22, no. 6, pp. 697–703, 2007.

[100] C. M. Powers, "Patellar kinematics, part ii: the influence of the depth of the trochlear groove in subjects with and without patellofemoral pain," *Physical therapy*, vol. 80, no. 10, pp. 965–973, 2000.

[101] C. M. Wille, R. L. Lenhart, S. Wang, D. G. Thelen, and B. C. Heiderscheit, "Ability of sagittal kinematic variables to estimate ground reaction forces and joint kinetics in running," *journal of orthopaedic & sports physical therapy*, vol. 44, no. 10, pp. 825–830, 2014.

[102] J. K. Loudon and M. P. Reiman, "Lower extremity kinematics in running athletes with and without a history of medial shin pain," *International journal of sports physical therapy*, vol. 7, no. 4, p. 356, 2012.

[103] D. Tiberio, "The effect of excessive subtalar joint pronation on patellofemoral mechanics: a theoretical model," *Journal of orthopaedic & Sports physical Therapy*, vol. 9, no. 4, pp. 160–165, 1987.

[104] B. Nohren, I. Davis, and J. Hamill, "Prospective study of the biomechanical factors associated with iliotibial band syndrome," *Clin Biomech*, vol. 22, pp. 951–956, 2007.

[105] D. M. Bazett-Jones, S. C. Cobb, W. E. Huddleston, K. M. O'CONNOR, B. S. Armstrong, and J. E. Earl-Boehm, "Effect of patellofemoral pain on strength and mechanics after an exhaustive run," *Medicine & Science in Sports & Exercise*, vol. 45, no. 7, pp. 1331–1339, 2013.

[106] S. Grau, I. Krauss, C. Maiwald, D. Axmann, T. Horstmann, and R. Best, "Kinematic classification of iliotibial band syndrome in runners," *Scandinavian journal of medicine & science in sports*, vol. 21, no. 2, pp. 184–189, 2011.

[107] M. Thompson, S. Lee, J. Seegmiller, and C. McGowan, "Kinematic and kinetic comparison of barefoot and shod running in mid/forefoot and rearfoot strike runners," *Gait & Posture*, vol. 41, no. 4, pp. 957–959, 2015.

[108] J. A. Mercer and S. Horsch, "Heel–toe running: A new look at the influence of foot strike pattern on impact force," *Journal of Exercise Science & Fitness*, vol. 13, no. 1, pp. 29–34, 2015.

[109] D. Kuhman, D. Melcher, and M. R. Paquette, "Ankle and knee kinetics between strike patterns at common training speeds in competitive male runners," *European Journal of Sport Science*, vol. 16, no. 4, pp. 433–440, 2016.

[110] C. A. Laughton, I. M. Davis, and J. Hamill, "Effect of strike pattern and orthotic intervention on tibial shock during running," *Journal of applied biomechanics*, vol. 19, no. 2, pp. 153–168, 2003.

[111] D. L. Goss and M. T. Gross, "A comparison of negative joint work and vertical ground reaction force loading rates in chi runners and rearfoot-striking runners," *journal of orthopaedic & sports physical therapy*, vol. 43, no. 10, pp. 685–692, 2013.

[112] J. R. Yong, A. Silder, K. L. Montgomery, M. Fredericson, and S. L. Delp, "Acute changes in foot strike pattern and cadence affect running parameters associated with tibial stress fractures," *Journal of biomechanics*, vol. 76, pp. 1–7, 2018.

[113] Y. Shih, K.-L. Lin, and T.-Y. Shiang, "Is the foot striking pattern more important than barefoot or shod conditions in running?," *Gait & posture*, vol. 38, no. 3, pp. 490–494, 2013.

[114] M. Ruder, S. Jamison, A. Tenforde, F. Mulloy, I. Davis, *et al.*, "Relationship of footstrike pattern and landing impacts during a marathon," *Medicine and science in sports and exercise*, vol. 51, no. 10, pp. 2073–2079, 2019.

[115] T. Delgado, E. Kubera-Shelton, R. Robb, R. Hickman, J. Dufek, and H. Wallmann, "Effects of foot strike on low back posture, shock attenuation, and comfort: Opo1165," *Journal of Orthopaedic & Sports Physical*, vol. 43, no. 1, 2013.

[116] K. G. Gerritsen, A. J. van den Bogert, and B. M. Nigg, "Direct dynamics simulation of the impact phase in heel-toe running," *Journal of biomechanics*, vol. 28, no. 6, pp. 661–668, 1995.

[117] J. P. Hall, C. Barton, P. R. Jones, and D. Morrissey, "The biomechanical differences between barefoot and shod distance running: a systematic review and preliminary meta-analysis," *Sports Medicine*, vol. 43, no. 12, pp. 1335–1353, 2013.

[118] C. Yeow, P. V. Lee, and J. C. Goh, "Regression relationships of landing height with ground reaction forces, knee flexion angles, angular velocities and joint powers during double-leg landing," *The Knee*, vol. 16, no. 5, pp. 381–386, 2009.

[119] B. P. Self and D. Paine, "Ankle biomechanics during four landing techniques.," *Medicine and science in sports and exercise*, vol. 33, no. 8, pp. 1338–1344, 2001.

[120] N. G. Elvin, A. A. Elvin, and S. P. Arnoczky, "Correlation between ground reaction force and tibial acceleration in vertical jumping," *Journal of applied biomechanics*, vol. 23, no. 3, pp. 180–189, 2007.

[121] M. A. Lafortune, E. M. Hennig, and M. J. Lake, "Dominant role of interface over knee angle for cushioning impact loading and regulating initial leg stiffness," *Journal of biomechanics*, vol. 29, no. 12, pp. 1523–1529, 1996.

[122] W. Potthast, G.-P. Brüggemann, A. Lundberg, and A. Arndt, "The influences of impact interface, muscle activity, and knee angle on impact forces and tibial and femoral accelerations occurring after external impacts," *Journal of applied biomechanics*, vol. 26, no. 1, pp. 1–9, 2010.

[123] W. B. Edwards, T. R. Derrick, and J. Hamill, "Musculoskeletal attenuation of impact shock in response to knee angle manipulation," *Journal of applied biomechanics*, vol. 28, no. 5, pp. 502–510, 2012.

[124] N. S. Romanov, *Dr. Nicholas Romanov's Pose Method of Running: A New Paradigm of Running.* Pose Tech Corp., 2002.

[125] D. Dreyer and K. Dreyer, *ChiRunning: A revolutionary approach to effortless, injury-free running.* Simon and Schuster, 2009.

[126] D. L. Goss and M. T. Gross, "A review of mechanics and injury trends among various running styles," 2012.

[127] D. E. Lieberman, M. Venkadesan, W. A. Werbel, A. I. Daoud, S. D'andrea, I. S. Davis, R. O. Mang'Eni, and Y. Pitsiladis, "Foot strike patterns and collision forces in habitually barefoot versus shod runners," *Nature*, vol. 463, no. 7280, pp. 531–535, 2010.

[128] N. Romanov and G. Fletcher, "Runners do not push off the ground but fall forwards via a gravitational torque," *Sports Biomechanics*, vol. 6, no. 3, pp. 434–452, 2007.

[129] R. E. Arendse, T. D. Noakes, L. B. Azevedo, N. Romanov, M. P. Schwellnus, and G. Fletcher, "Reduced eccentric loading of the knee with the pose running method," *Medicine & Science in Sports & Exercise*, 2004.

[130] G. Fletcher, R. Bartlett, N. Romanov, and A. Fotouhi, "Pose® method technique improves running performance without economy changes," *International Journal of Sports Science & Coaching*, vol. 3, no. 3, pp. 365–380, 2008.

[131] R. X. Wei, I. P. Au, F. O. Lau, J. H. Zhang, Z. Y. Chan, A. J. MacPhail, A. L. Mangubat, G. Pun, and R. T. Cheung, "Running biomechanics before and after pose® method gait retraining in distance runners," *Sports biomechanics*, pp. 1–16, 2019.

[132] I. Buist, S. W. Bredeweg, B. Bessem, W. Van Mechelen, K. A. Lemmink, and R. L. Diercks, "Incidence and risk factors of running-related injuries during preparation for a 4-mile recreational running event," *British journal of sports medicine*, vol. 44, no. 8, pp. 598–604, 2010.

[133] A. D. Lopes, L. O. P. Costa, B. T. Saragiotto, T. P. Yamato, F. Adami, and E. Verhagen, "Musculoskeletal pain is prevalent among recreational runners who are about to compete: an observational study of 1049 runners," *Journal of physiotherapy*, vol. 57, no. 3, pp. 179–182, 2011.

[134] L. Linton and S. Valentin, "Running with injury: A study of uk novice and recreational runners and factors associated with running related injury," *Journal of science and medicine in sport*, vol. 21, no. 12, pp. 1221–1225, 2018.

[135] P. Satterthwaite, R. Norton, P. Larmer, and E. Robinson, "Risk factors for injuries and other health problems sustained in a marathon.," *British Journal of Sports Medicine*, vol. 33, no. 1, pp. 22–26, 1999.

[136] J. Taunton, M. Ryan, D. Clement, D. McKenzie, D. Lloyd-Smith, and B. Zumbo, "A prospective study of running injuries: the vancouver sun run "in training" clinics," *British journal of sports medicine*, vol. 37, no. 3, pp. 239–244, 2003.

[137] B. Kluitenberg, M. van Middelkoop, D.-W. Smits, E. Verhagen, F. Hartgens, R. Diercks, and H. van der Worp, "The nl start2run study: Incidence and risk factors of running-related injuries in novice runners," *Scandinavian journal of medicine & science in sports*, vol. 25, no. 5, pp. e515–e523, 2015.

[138] J. M. Hootman, C. A. Macera, B. E. Ainsworth, M. Martin, C. L. Addy, and S. N. Blair, "Predictors of lower extremity injury among recreationally active adults," *Clinical Journal of Sport Medicine*, vol. 12, no. 2, pp. 99–106, 2002.

[139] C. H. Rasmussen, R. O. Nielsen, M. S. Juul, and S. Rasmussen, "Weekly running volume and risk of running-related injuries among marathon runners," *International journal of sports physical therapy*, vol. 8, no. 2, p. 111, 2013.

[140] A. Hirschmüller, V. Frey, L. Konstantinidis, H. Baur, H.-H. Dickhuth, N. P. Suedkamp, and P. Helwig, "Prognostic value of achilles tendon doppler sonography in asymptomatic runners.," *Medicine and science in sports and exercise*, vol. 44, no. 2, pp. 199–205, 2012.

[141] J. L. Kelsey, L. K. Bachrach, E. Procter-Gray, J. Nieves, G. A. Greendale, M. Sowers, B. W. Brown Jr, K. A. Matheson, S. L. Crawford, and K. L. Cobb, "Risk factors for stress fracture among young female cross-country runners," *Medicine & Science in Sports & Exercise*, vol. 39, no. 9, pp. 1457–1463, 2007.

[142] R. K. Fukuchi, D. J. Stefanyshyn, L. Stirling, M. Duarte, and R. Ferber, "Flexibility, muscle strength and running biomechanical adaptations in older runners," *Clinical Biomechanics*, vol. 29, no. 3, pp. 304–310, 2014.

[143] W. Van Mechelen, "Running injuries," *Sports medicine*, vol. 14, no. 5, pp. 320–335, 1992.

[144] B. Marti, J. P. Vader, C. E. Minder, and T. Abelin, "On the epidemiology of running injuries: the 1984 bern grand-prix study," *The American Journal of Sports Medicine*, vol. 16, no. 3, pp. 285–294, 1988.

[145] M. Van Middelkoop, J. Kolkman, J. Van Ochten, S. Bierma-Zeinstra, and B. W. Koes, "Risk factors for lower extremity injuries among male marathon runners," *Scandinavian journal of medicine & science in sports*, vol. 18, no. 6, pp. 691–697, 2008.

[146] M. Besomi, J. Leppe, M. V. Mauri-Stecca, T. L. Hooper, and P. S. Sizer, "Training volume and previous injury as associated factors for running-related injuries by race distance: a cross-sectional study," 2019.

[147] F. Porciuncula, A. V. Roto, D. Kumar, I. Davis, S. Roy, C. J. Walsh, and L. N. Awad, "Wearable movement sensors for rehabilitation: a focused review of technological and clinical advances," *Pm&r*, vol. 10, no. 9, pp. S220–S232, 2018.

[148] B. T. Saragiotto, T. P. Yamato, L. C. H. Junior, M. J. Rainbow, I. S. Davis, and A. D. Lopes, "What are the main risk factors for running-related injuries?," *Sports medicine*, vol. 44, no. 8, pp. 1153–1163, 2014.

[149] J. Fulton, K. Wright, M. Kelly, B. Zebrosky, M. Zanis, C. Drvol, and R. Butler, "Injury risk is altered by previous injury: a systematic review of the literature and presentation of causative neuromuscular factors," *International journal of sports physical therapy*, vol. 9, no. 5, p. 583, 2014.

[150] I. Buist, S. W. Bredeweg, K. A. Lemmink, W. Van Mechelen, and R. L. Diercks, "Predictors of running-related injuries in novice runners enrolled in a systematic training program: a prospective cohort study," *The American journal of sports medicine*, vol. 38, no. 2, pp. 273–280, 2010.

[151] R. Nielsen, I. Buist, H. Sørensen, M. Lind, and S. Rasmussen, "Running-related injuries among novice runners: a 1-year prospective follow-up study," in *Sports Medicine Congress, 31 January-2 February 2013, Kolding, Denmark:: From Research to Clinical Practice*, p. 27, 2013.

[152] C. A. Macera, R. R. Pate, K. E. Powell, K. L. Jackson, J. S. Kendrick, and T. E. Craven, "Predicting lower-extremity injuries among habitual runners," *Archives of internal medicine*, vol. 149, no. 11, pp. 2565–2568, 1989.

[153] B. Van Hooren, J. T. Fuller, J. D. Buckley, J. R. Miller, K. Sewell, G. Rao, C. Barton, C. Bishop, and R. W. Willy, "Is motorized treadmill running biomechanically comparable to overground

running? a systematic review and meta-analysis of cross-over studies," *Sports medicine*, vol. 50, no. 4, pp. 785–813, 2020.

[154] F. Di Caprio, R. Buda, M. Mosca, A. Calabrò, and S. Giannini, "Foot and lower limb diseases in runners: assessment of risk factors," *Journal of sports science & medicine*, vol. 9, no. 4, p. 587, 2010.

[155] S. P. Messier, D. G. Edwards, D. F. Martin, R. B. Lowery, D. W. Cannon, M. K. James, W. W. Curl, H. M. Read Jr, and D. M. Hunter, "Etiology of iliotibial band friction syndrome in distance runners.," *Medicine and science in sports and exercise*, vol. 27, no. 7, pp. 951–960, 1995.

[156] A. C. Ferreira, J. M. C. Dias, R. d. M. Fernandes, G. S. Sabino, M. T. S. d. Anjos, and D. C. Felício, "Prevalence and asociated risks of injury in amateur street runners from belo horizonte, mg," *Revista Brasileira de Medicina do Esporte*, vol. 18, no. 4, pp. 252–255, 2012.

[157] T. J. Gabbett, "The training—injury prevention paradox: should athletes be training smarter and harder?," *British journal of sports medicine*, vol. 50, no. 5, pp. 273–280, 2016.

[158] L. C. H. Junior, L. O. P. Costa, and A. D. Lopes, "Previous injuries and some training characteristics predict running-related injuries in recreational runners: a prospective cohort study," *Journal of Physiotherapy*, vol. 59, no. 4, pp. 263–269, 2013.

[159] L. Malisoux, J. Ramesh, R. Mann, R. Seil, A. Urhausen, and D. Theisen, "Can parallel use of different running shoes decrease running-related injury risk?," *Scandinavian journal of medicine & science in sports*, vol. 25, no. 1, pp. 110–115, 2015.

[160] K. A. McKean, N. A. Manson, and W. D. Stanish, "Musculoskeletal injury in the masters runners," *Clinical Journal of Sport Medicine*, vol. 16, no. 2, pp. 149–154, 2006.

[161] D. Theisen, L. Malisoux, J. Genin, N. Delattre, R. Seil, and A. Urhausen, "Influence of midsole hardness of standard cushioned shoes on running-related injury risk," *British Journal of Sports Medicine*, vol. 48, no. 5, pp. 371–376, 2014.

[162] J. P. Amaro and S. Patrão, "A survey of sensor fusion algorithms for sport and health monitoring applications," in *IECON 2016-42nd Annual Conference of the IEEE Industrial Electronics Society*, pp. 5171–5176, IEEE, 2016.

[163] Z. M. M. Htun, M. M. Latt, C. M. Nwe, and S. S. Y. Mon, "performance comparison of experimental-based kalman filter and complementary filter for imu sensor fusion by applying quadrature encoder," *International Journal of Scientific and Research Publications*, vol. 8, no. 11, 2018.

[164] C. Mazza, M. Donati, J. McCamley, P. Picerno, and A. Cappozzo, "An optimized kalman filter for the estimate of trunk orientation from inertial sensors data during treadmill walking," *Gait & posture*, vol. 35, no. 1, pp. 138–142, 2012.

[165] G. Ligorio and A. M. Sabatini, "A novel kalman filter for human motion tracking with an inertial-based dynamic inclinometer," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 8, pp. 2033–2043, 2015.

[166] E. Foxlin, "Inertial head-tracker sensor fusion by a complementary separate-bias kalman filter," in *Proceedings of the IEEE 1996 Virtual Reality Annual International Symposium*, pp. 185–194, IEEE, 1996.

[167] H. J. Luinge and P. H. Veltink, "Measuring orientation of human body segments using miniature gyroscopes and accelerometers," *Medical and Biological Engineering and computing*, vol. 43, no. 2, pp. 273–282, 2005.

[168] D. Roetenberg, *Inertial and magnetic sensing of human motion*. These de doctorat, 2006.

[169] W. Hu, E. Charry, M. Umer, A. Ronchi, and S. Taylor, "An inertial sensor system for measurements of tibia angle with applications to knee valgus/varus detection," in *2014 IEEE Ninth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP)*, pp. 1–6, IEEE, 2014.

[170] R. E. Kalman, "A new approach to linear filtering and prediction problems," 1960.

[171] F. L. Lewis, L. Xie, and D. Popa, *Optimal and robust estimation: with an introduction to stochastic control theory*. CRC press, 2017.

[172] Z.-Q. Zhang, X.-L. Meng, and J.-K. Wu, "Quaternion-based kalman filter with vector selection for accurate orientation tracking," *IEEE Transactions on Instrumentation and Measurement*, vol. 61, no. 10, pp. 2817–2824, 2012.

[173] J. K. Lee, E. J. Park, and S. N. Robinovitch, "Estimation of attitude and external acceleration using inertial sensor measurement during various dynamic conditions," *IEEE transactions on instrumentation and measurement*, vol. 61, no. 8, pp. 2262–2273, 2012.

[174] R. Zhu and Z. Zhou, "A real-time articulated human motion tracking using tri-axis inertial/magnetic sensors package," *IEEE Transactions on Neural systems and rehabilitation engineering*, vol. 12, no. 2, pp. 295–302, 2004.

[175] G. Cooper, I. Sheret, L. McMillian, K. Siliverdis, N. Sha, D. Hodgins, L. Kenney, and D. Howard, "Inertial sensor-based knee flexion/extension angle estimation," *Journal of biomechanics*, vol. 42, no. 8, pp. 2678–2685, 2009.

[176] C. Jakob, P. Kugler, F. Hebenstreit, S. Reinfelder, U. Jensen, D. Schuldhaus, M. Lochmann, and B. M. Eskofier, "Estimation of the knee flexion-extension angle during dynamic sport motions using body-worn inertial sensors," in *Proceedings of the 8th International Conference on Body Area Networks*, pp. 289–295, 2013.

[177] A. M. Sabatini, "Quaternion-based extended kalman filter for determining orientation by inertial and magnetic sensing," *IEEE transactions on Biomedical Engineering*, vol. 53, no. 7, pp. 1346–1356, 2006.

[178] A. M. Sabatini, "Kalman-filter-based orientation determination using inertial/magnetic sensors: Observability analysis and performance evaluation," *Sensors*, vol. 11, no. 10, pp. 9182–9206, 2011.

[179] D. R. Labbé, D. Li, G. Grimard, J. A. de Guise, and N. Hagemeister, "Quantitative pivot shift assessment using combined inertial and magnetic sensing," *Knee Surgery, Sports Traumatology, Arthroscopy*, vol. 23, no. 8, pp. 2330–2338, 2015.

[180] X. Yun and E. R. Bachmann, "Design, implementation, and experimental results of a quaternion-based kalman filter for human body motion tracking," *IEEE transactions on Robotics*, vol. 22, no. 6, pp. 1216–1227, 2006.

[181] S. J. Julier and J. K. Uhlmann, "A new extension of the kalman filter to nonlinear systems," in *Signal processing, sensor fusion, and target recognition VI*, vol. 3068, pp. 182–193, International Society for Optics and Photonics, 1997.

[182] E. A. Wan and R. Van Der Merwe, "The unscented kalman filter for nonlinear estimation," in *Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (Cat. No. 00EX373)*, pp. 153–158, Ieee, 2000.

[183] M. El-Gohary and J. McNames, "Shoulder and elbow joint angle tracking with inertial sensors," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 9, pp. 2635–2641, 2012.

[184] J. K. Uhlmann, "Simultaneous map building and localization for real time applications," tech. rep., Technical report, University of Oxford, 1994. Transfer thesis, 1994.

[185] M. H. Ebrahim, J. M. Feldman, and I. Bar-Kana, "A robust sensor fusion method for heart rate estimation," *Journal of clinical monitoring*, vol. 13, no. 6, pp. 385–393, 1997.

[186] M. S. Karunarathne, S. W. Ekanayake, and P. N. Pathirana, "An adaptive complementary filter for inertial sensor based data fusion to track upper body motion," in *7th International Conference on Information and Automation for Sustainability*, pp. 1–5, IEEE, 2014.

[187] V. Kubelka and M. Reinstein, "Complementary filtering approach to orientation estimation using inertial sensors only," in *2012 IEEE international conference on robotics and automation*, pp. 599–605, IEEE, 2012.

[188] O. Sarbishei, "On the accuracy improvement of low-power orientation filters using imu and marg sensor arrays," in *2016 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 1542–1545, IEEE, 2016.

[189] H. Fourati, N. Manamanni, L. Afilal, and Y. Handrich, "Complementary observer for body segments motion capturing by inertial and magnetic sensors," *IEEE/ASME transactions on Mechatronics*, vol. 19, no. 1, pp. 149–157, 2012.

[190] B. Fan, Q. Li, C. Wang, and T. Liu, "An adaptive orientation estimation method for magnetic and inertial sensors in the presence of magnetic disturbances," *Sensors*, vol. 17, no. 5, p. 1161, 2017.

[191] Y. Tian, H. Wei, and J. Tan, "An adaptive-gain complementary filter for real-time human motion tracking with marg sensors in free-living environments," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 21, no. 2, pp. 254–264, 2012.

[192] Y. Tian and J. Tan, "A fast adaptive-gain orientation filter of inertial/magnetic data for human motion tracking in free-living environments," in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 6760–6763, IEEE, 2012.

[193] M. Rasteiro, H. Costelha, L. Bento, and P. Assuncao, "Accuracy versus complexity of marg-based filters for remote control pointing devices," in *2015 IEEE International Conference on Consumer Electronics-Taiwan*, pp. 51–52, IEEE, 2015.

[194] M. Rasteiro, H. Costelha, L. Bento, M. Barata, and P. Assunçao, "Low-complexity marg algorithms for increased accuracy in space pointing devices," in *2015 IEEE 1st International Workshop on Consumer Electronics (CE WS)*, pp. 87–90, IEEE, 2015.

[195] J. Cockcroft, J. H. Muller, and C. Scheffer, "A novel complimentary filter for tracking hip angles during cycling using wireless inertial sensors and dynamic acceleration estimation," *IEEE Sensors Journal*, vol. 14, pp. 2864–2871, 2014.

[196] Y. Huang, W. Jirattigalachote, M. R. Cutkosky, X. Zhu, and P. B. Shull, "Novel foot progression angle algorithm estimation via foot-worn, magneto-inertial sensing," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 11, pp. 2278–2285, 2016.

[197] K. Kim, J. S. Kim, and Y. J. Kim, "Application of nonlinear complementary filters to human motion analysis," in *2015 17th International Conference on E-health Networking, Application & services (HealthCom)*, pp. 594–595, IEEE, 2015.

[198] O. Sarbishei, "A platform and methodology enabling real-time motion pattern recognition on low-power smart devices," in *2019 IEEE 5th World Forum on Internet of Things (WF-IoT)*, pp. 269–272, IEEE, 2019.

[199] A. D. Young, "Comparison of orientation filter algorithms for realtime wireless inertial posture tracking," in *2009 Sixth International Workshop on Wearable and Implantable Body Sensor Networks*, pp. 59–64, IEEE, 2009.

[200] J. R. Wertz, *Spacecraft attitude determination and control*, vol. 73. Springer Science & Business Media, 2012.

[201] M. D. Shuster and S. D. Oh, "Three-axis attitude determination from vector observations," *Journal of guidance and Control*, vol. 4, no. 1, pp. 70–77, 1981.

[202] X. Yun, E. R. Bachmann, and R. B. McGhee, "A simplified quaternion-based algorithm for orientation estimation from earth gravity and magnetic field measurements," *IEEE Transactions on instrumentation and measurement*, vol. 57, no. 3, pp. 638–650, 2008.

[203] G. Wahba, "Least squares estimate of spacecraft attitude," *SIAM Review*, vol. 7, no. 3, 1965.

[204] J. L. Crassidis, F. L. Markley, and Y. Cheng, "Survey of nonlinear attitude estimation methods," *Journal of guidance, control, and dynamics*, vol. 30, no. 1, pp. 12–28, 2007.

[205] R. G. Valenti, I. Dryanovski, and J. Xiao, "Keeping a good attitude: A quaternion-based orientation filter for imus and margs," *Sensors*, vol. 15, no. 8, pp. 19302–19330, 2015.

[206] M. D. Shuster, "Kalman filtering of spacecraft attitude and the quest model," *Journal of the Astronautical Sciences*, vol. 38, pp. 377–393, 1990.

[207] M. Rhudy, Y. Gu, J. Gross, S. Gururajan, and M. R. Napolitano, "Sensitivity analysis of extended and unscented kalman filters for attitude estimation," *Journal of Aerospace Information Systems*, vol. 10, no. 3, pp. 131–143, 2013.

[208] J. J. LaViola, "A comparison of unscented and extended kalman filtering for estimating quaternion motion," in *Proceedings of the 2003 American Control Conference, 2003.*, vol. 3, pp. 2435–2440, IEEE, 2003.

[209] F. Öhberg, R. Lundström, and H. Grip, "Comparative analysis of different adaptive filters for tracking lower segments of a human body using inertial motion sensors," *Measurement science and technology*, vol. 24, no. 8, p. 085703, 2013.

[210] M. Giurato and M. Lovera, "Quadrotor attitude determination: a comparison study," in *2016 IEEE Conference on Control Applications (CCA)*, pp. 21–26, IEEE, 2016.

[211] L. S. V. Valencia, F. B. Schneider, A. Leal-Junior, P. Caicedo-Rodrguez, W. A. Sierra-Arvalo, L. E. Rodrguez-Cheu, T. F. Bastos-Filho, and A. Frizera, "Sleeve for knee angle monitoring: An imu-pof sensor fusion system," *IEEE journal of biomedical and health informatics*, 2020.

[212] R. Mahony, T. Hamel, and J.-M. Pflimlin, "Nonlinear complementary filters on the special orthogonal group," *IEEE Transactions on automatic control*, vol. 53, no. 5, pp. 1203–1218, 2008.

[213] P. Daponte, L. De Vito, M. Riccio, and C. Sementa, "Experimental comparison of orientation estimation algorithms in motion tracking for rehabilitation," in *2014 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, pp. 1–6, IEEE, 2014.

[214] A. Cavallo, A. Cirillo, P. Cirillo, G. De Maria, P. Falco, C. Natale, and S. Pirozzi, "Experimental comparison of sensor fusion algorithms for attitude estimation," *IFAC Proceedings Volumes*, vol. 47, no. 3, pp. 7585–7591, 2014.

[215] H.-P. Brückner, B. Krüger, and H. Blume, "Reliable orientation estimation for mobile motion capturing in medical rehabilitation sessions based on inertial measurement units," *Microelectronics Journal*, vol. 45, no. 12, pp. 1603–1611, 2014.

[216] D. Parikh, S. Vohra, and M. Kaveshgar, "Comparison of attitude estimation algorithms with imu under external acceleration," in *2021 IEEE International Symposium on Smart Electronic Systems (iSES)(Formerly iNiS)*, pp. 123–126, IEEE, 2021.

[217] W. B. Edwards, D. Taylor, T. J. Rudolphi, J. C. Gillette, and T. R. Derrick, "Effects of running speed on a probabilistic stress fracture model," *Clinical Biomechanics*, vol. 25, no. 4, pp. 372–377, 2010.

[218] D. A. Sharma, M. F. Chevidikunnan, F. R. Khan, and R. A. Gaowgzeh, "Effectiveness of knowledge of result and knowledge of performance in the learning of a skilled motor activity by healthy young adults," *Journal of physical therapy science*, vol. 28, no. 5, pp. 1482–1486, 2016.

[219] R. Magill and D. Anderson, *Motor learning and control*. McGraw-Hill Publishing New York, 2010.

[220] M. Zubiaur, A. Ona, and J. Delgado, "Learning volleyball serves: a preliminary study of the effects of knowledge of performance and of results," *Perceptual and motor skills*, vol. 89, no. 1, pp. 223–232, 1999.

[221] S. M. Henry and D. S. Teyhen *journal of orthopaedic & sports physical therapy*, vol. 37, no. 10, pp. 627–634, 2007.

[222] D. Sharma, "Knowledge of result versus knowledge of performance in learning motor skilled activity for upper limb in recovery stage of stroke–a randomized experimental study," *Medical Science*, vol. 5, no. 9, 2016.

[223] B. Van Hooren, J. Goudsmit, J. Restrepo, and S. Vos, "Real-time feedback by wearables in running: Current approaches, challenges and suggestions for improvements," *Journal of Sports Sciences*, vol. 38, no. 2, pp. 214–230, 2020.

[224] R. T. Cheung and I. S. Davis, "Landing pattern modification to improve patellofemoral pain in runners: a case series," *Journal of orthopaedic & sports physical therapy*, vol. 41, no. 12, pp. 914–919, 2011.

[225] B. Noehren, J. Scholz, and I. Davis, "The effect of real-time gait retraining on hip kinematics, pain and function in subjects with patellofemoral pain syndrome," *British journal of sports medicine*, vol. 45, no. 9, pp. 691–696, 2011.

[226] H. P. Crowell, C. E. Milner, J. Hamill, and I. S. Davis, "Reducing impact loading during running with the use of real-time visual feedback," *journal of orthopaedic & sports physical therapy*, vol. 40, no. 4, pp. 206–213, 2010.

[227] M. Eriksson, K. A. Halvorsen, and L. Gullstrand, "Immediate effect of visual and auditory feedback to control the running mechanics of well-trained athletes," *Journal of sports sciences*, vol. 29, no. 3, pp. 253–262, 2011.

[228] J. J. Tate and C. E. Milner, "Sound-intensity feedback during running reduces loading rates and impact peak," *journal of orthopaedic & sports physical therapy*, vol. 47, no. 8, pp. 565–569, 2017.

[229] A. C. Clansey, M. Hanlon, E. S. Wallace, A. Nevill, and M. J. Lake, "Influence of tibial shock feedback training on impact loading and running economy.," *Medicine and science in sports and exercise*, vol. 46, no. 5, pp. 973–981, 2014.

[230] A. Garofolini, L. Oppici, and S. Taylor, "A real-time feedback method to reduce loading rate during running: Effect of combining direct and indirect feedback," *Journal of Sports Sciences*, vol. 38, no. 21, pp. 2446–2453, 2020.

[231] S. Rothstein, "Efficacy of an audio-based biofeedback intervention to modify running gait in female runners," 2020.

[232] X. Phan, T. L. Grisbrook, K. Wernli, S. M. Stearne, P. Davey, and L. Ng, "Running quietly reduces ground reaction force and vertical loading rate and alters foot strike technique," *Journal of sports sciences*, vol. 35, no. 16, pp. 1636–1642, 2017.

[233] P. Van den Berghe, V. Lorenzoni, J. Gerlo, B. Breine, R. Derie, J. Six, M. Leman, and D. De Clercq, "Real-time music-based biofeedback to reduce impact loading during over-ground running," in *42nd Annual meeting of the American Society of Biomechanics (ASB)*, 2018.

[234] P. Van den Berghe, V. Lorenzoni, R. Derie, J. Six, J. Gerlo, M. Leman, and D. De Clercq, "Music-based biofeedback to reduce tibial shock in over-ground running: a proof-of-concept study," *Scientific reports*, vol. 11, no. 1, pp. 1–12, 2021.

[235] P. Van den Berghe, R. Derie, P. Bauwens, J. Gerlo, V. Segers, M. Leman, and D. De Clercq, "Reducing the peak tibial acceleration of running by music-based biofeedback: A quasi-randomized controlled trial," *Scandinavian Journal of Medicine & Science in Sports*, 2022.

[236] M. W. Creaby and M. M. F. Smith, "Retraining running gait to reduce tibial loads with clinician or accelerometry guided feedback," *Journal of science and medicine in sport*, vol. 19, no. 4, pp. 288–292, 2016.

[237] E. Gray, M. Sweeney, M. Creaby, and M. Smith, "Gait retraining using visual and verbal feedback in runners," in *ISBS-Conference Proceedings Archive*, 2012.

[238] J. H. Zhang, Z. Y. Chan, I. P. Au, W. W. An, and R. T. Cheung, "Can runners maintain a newly learned gait pattern outside a laboratory environment following gait retraining?," *Gait & posture*, vol. 69, pp. 8–12, 2019.

[239] C. M. Wood and K. Kipp, "Use of audio biofeedback to reduce tibial impact accelerations during running," *Journal of biomechanics*, vol. 47, no. 7, pp. 1739–1741, 2014.

[240] K. R. Sheerin, D. Reid, D. Taylor, and T. F. Besier, "The effectiveness of real-time haptic feedback gait retraining for reducing resultant tibial acceleration with runners," *Physical Therapy in Sport*, vol. 43, pp. 173–180, 2020.

[241] E. Ching, W. W.-K. An, I. P. H. Au, J. H. Zhang, Z. Y. Chan, G. Shum, and R. T. Cheung, "Impact loading during distracted running before and after auditory gait retraining," *International journal of sports medicine*, vol. 39, no. 14, pp. 1075–1080, 2018.

[242] R. T. Cheung, W. W. An, I. P. Au, J. H. Zhang, Z. Y. Chan, and A. J. MacPhail, "Control of impact loading during distracted running before and after gait retraining in runners," *Journal of sports sciences*, vol. 36, no. 13, pp. 1497–1501, 2018.

[243] M. K. Shah and G. W. Stewart, "Sacral stress fractures: an unusual cause of low back pain in an athlete," *Spine*, vol. 27, no. 4, pp. E104–E108, 2002.

[244] Y. Huang, H. Xia, G. Chen, S. Cheng, R. T. Cheung, and P. B. Shull, "Foot strike pattern, step rate, and trunk posture combined gait modifications to reduce impact loading during running," *Journal of biomechanics*, vol. 86, pp. 102–109, 2019.

[245] L. Hunter, Q. A. Louw, and S.-M. van Niekerk, "Effect of running retraining on pain, function, and lower-extremity biomechanics in a female runner with iliotibial band syndrome," *Journal of sport rehabilitation*, vol. 23, no. 2, pp. 145–157, 2014.

[246] R. W. Willy, J. P. Scholz, and I. S. Davis, "Mirror gait retraining for the treatment of patellofemoral pain in female runners," *Clinical Biomechanics*, vol. 27, no. 10, pp. 1045–1051, 2012.

[247] J. L. Roper, E. M. Harding, D. Doerfler, J. G. Dexter, L. Kravitz, J. S. Dufek, and C. M. Mermier, "The effects of gait retraining in runners with patellofemoral pain: a randomized trial," *Clinical biomechanics*, vol. 35, pp. 14–22, 2016.

[248] R. Willy, L. Buchenic, K. Rogacki, J. Ackerman, A. Schmidt, and J. Willson, "In-field gait retraining and mobile monitoring to address running biomechanics associated with tibial stress fracture," *Scandinavian journal of medicine & science in sports*, vol. 26, no. 2, pp. 197–205, 2016.

[249] E. M. Miller, M. S. Crowell, J. B. Morris, J. S. Mason, R. Zifchock, and D. L. Goss, "Gait retraining improves running impact loading and function in previously injured us military cadets: A pilot study," *Military Medicine*, vol. 186, no. 11-12, pp. e1077–e1087, 2021.

[250] H. Argunsah Bayram and B. Yalcin, "The influence of biofeedback on physiological and kinematic variables of treadmill running," *International Journal of Performance Analysis in Sport*, vol. 21, no. 1, pp. 156–169, 2021.

[251] V. Lorenzoni, P.-J. Maes, P. Van den Berghe, D. De Clercq, T. De Bie, and M. Leman, "A biofeedback music-sonification system for gait retraining," in *Proceedings of the 5th International Conference on Movement and Computing*, pp. 1–5, 2018.

[252] J. Spencer, S. L. Wolf, and T. M. Kesar, "Biofeedback for post-stroke gait retraining: A review of current evidence and future research directions in the context of emerging technologies," *Frontiers in Neurology*, vol. 12, p. 419, 2021.

[253] C. Napier, J.-F. Esculier, and M. A. Hunt, "Gait retraining: out of the lab and onto the streets with the benefit of wearables," 2017.

[254] A. Sandryhaila and J. M. Moura, "Big data analysis with signal processing on graphs: Representation and processing of massive data sets with irregular structure," *IEEE signal processing magazine*, vol. 31, no. 5, pp. 80–90, 2014.

[255] M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," *Science*, vol. 349, no. 6245, pp. 255–260, 2015.

[256] J. Qiu, Q. Wu, G. Ding, Y. Xu, and S. Feng, "A survey of machine learning for big data processing," *EURASIP Journal on Advances in Signal Processing*, vol. 2016, no. 1, pp. 1–16, 2016.

[257] C. M. Bishop, "Pattern recognition," *Machine learning*, vol. 128, no. 9, 2006.

[258] N. Burkart and M. F. Huber, "A survey on the explainability of supervised machine learning," *Journal of Artificial Intelligence Research*, vol. 70, pp. 245–317, 2021.

[259] B. C. Luu, A. L. Wright, H. S. Haeberle, J. M. Karnuta, M. S. Schickendantz, E. C. Makhni, B. U. Nwachukwu, R. J. Williams III, and P. N. Ramkumar, "Machine learning outperforms logistic regression analysis to predict next-season nhl player injury: an analysis of 2322 players from 2007 to 2017," *Orthopaedic journal of sports medicine*, vol. 8, no. 9, p. 2325967120953404, 2020.

[260] H. Van Eetvelde, L. D. Mendonça, C. Ley, R. Seil, and T. Tischer, "Machine learning methods in sport injury prediction and prevention: a systematic review," *Journal of experimental orthopaedics*, vol. 8, no. 1, pp. 1–15, 2021.

[261] E. Rovini, C. Maremmani, A. Moschetti, D. Esposito, and F. Cavallo, "Comparative motor pre-clinical assessment in parkinson's disease using supervised machine learning approaches," *Annals of biomedical engineering*, vol. 46, no. 12, pp. 2057–2068, 2018.

[262] F. Ayala, A. López-Valenciano, J. A. G. Martín, M. D. S. Croix, F. J. Vera-Garcia, M. del Pilar García-Vaquero, I. Ruiz-Pérez, and G. D. Myer, "A preventive model for hamstring injuries in professional soccer: learning algorithms," *International journal of sports medicine*, vol. 40, no. 05, pp. 344–353, 2019.

[263] V. Mandalapu, N. Homdee, J. M. Hart, J. Lach, S. Bodkin, and J. Gong, "Developing computational models for personalized acl injury classification," in *2019 IEEE 16th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, pp. 1–4, IEEE, 2019.

[264] C. Richter, E. Petushek, H. Grindem, A. Franklyn-Miller, R. Bahr, and T. Krosshaug, "Cross-validation of a machine learning algorithm that determines anterior cruciate ligament rehabilitation status and evaluation of its ability to predict future injury," *Sports biomechanics*, pp. 1–11, 2021.

[265] P. H. Borgstrom, K. L. Markolf, Y. Wang, X. Xu, P. R. Yang, N. B. Joshi, M. G. Yeranosian, F. A. Petrigliano, S. L. Hame, W. J. Kaiser, *et al.*, "Use of inertial sensors to predict pivot-shift grade and diagnose an acl injury during preoperative testing," *The American journal of sports medicine*, vol. 43, no. 4, pp. 857–864, 2015.

[266] W. Wu, W. Zeng, L. Ma, C. Yuan, and Y. Zhang, "Modeling and classification of gait patterns between anterior cruciate ligament deficient and intact knees based on phase space reconstruction, euclidean distance and neural networks," *Biomedical engineering online*, vol. 17, no. 1, pp. 1–19, 2018.

[267] C. Richter, E. King, S. Strike, and A. Franklyn-Miller, "Objective classification and scoring of movement deficiencies in patients with anterior cruciate ligament reconstruction," *PloS one*, vol. 14, no. 7, p. e0206024, 2019.

[268] S. Tedesco, C. Crowe, A. Ryan, M. Sica, S. Scheurer, A. M. Clifford, K. N. Brown, and B. O'Flynn, "Motion sensors-based machine learning approach for the identification of anterior cruciate ligament gait patterns in on-the-field activities in rugby players," *Sensors*, vol. 20, no. 11, p. 3029, 2020.

[269] J. Taborri, L. Molinaro, A. Santospagnuolo, M. Vetrano, M. C. Vulpiani, and S. Rossi, "A machine-learning approach to measure the anterior cruciate ligament injury risk in female basketball players," *Sensors*, vol. 21, no. 9, p. 3141, 2021.

[270] S. Jauhiainen, J.-P. Kauppi, M. Leppänen, K. Pasanen, J. Parkkari, T. Vasankari, P. Kannus, and S. Äyrämö, "New machine learning approach for detection of injury risk factors in young team sport athletes," *International journal of sports medicine*, vol. 42, no. 02, pp. 175–182, 2021.

[271] M. Nazarahari, K. M. Chan, and H. Rouhani, "A novel instrumented shoulder functional test using wearable sensors in patients with brachial plexus injury," *Journal of Shoulder and Elbow Surgery*, vol. 30, no. 8, pp. e493–e502, 2021.

[272] S. Gore, *The Identification of Foot-Strike Patterns and Prediction of Running Related Injuries*. PhD thesis, Dublin, National College of Ireland, 2020.

[273] J. Martínez-Gramage, J. P. Albiach, I. N. Moltó, J. J. Amer-Cuenca, V. Huesa Moreno, and E. Segura-Ortí, "A random forest machine learning framework to reduce running injuries in young triathletes," *Sensors*, vol. 20, no. 21, p. 6388, 2020.

[274] M. Henriquez, J. Sumner, M. Faherty, T. Sell, and B. Bent, "Machine learning to predict lower extremity musculoskeletal injury risk in student athletes," *Frontiers in sports and active living*, vol. 2, 2020.

[275] A. Rossi, L. Pappalardo, P. Cintia, J. Fernández, M. F. Iaia, and D. Medina, "Who is going to get hurt? predicting injuries in professional soccer.," in *MLSA@ PKDD/ECML*, pp. 21–30, 2017.

[276] A. López-Valenciano, F. Ayala, J. M. Puerta, M. D. S. Croix, F. Vera-García, S. Hernández-Sánchez, I. Ruiz-Pérez, and G. Myer, "A preventive model for muscle injuries: a novel approach based on learning algorithms," *Medicine and science in sports and exercise*, vol. 50, no. 5, p. 915, 2018.

[277] J. L. Oliver, F. Ayala, M. B. D. S. Croix, R. S. Lloyd, G. D. Myer, and P. J. Read, "Using machine learning to improve our understanding of injury risk and prediction in elite male youth football players," *Journal of science and medicine in sport*, vol. 23, no. 11, pp. 1044–1048, 2020.

[278] I. Ruiz-Pérez, A. López-Valenciano, S. Hernández-Sánchez, J. M. Puerta-Callejón, M. De Ste Croix, P. Sainz de Baranda, and F. Ayala, "A field-based approach to determine soft tissue injury risk in elite futsal using novel machine learning techniques," *Frontiers in psychology*, vol. 12, p. 195, 2021.

[279] D. L. Carey, K.-L. Ong, R. Whiteley, K. M. Crossley, J. Crow, and M. E. Morris, "Predictive modelling of training loads and injury in australian football," *arXiv preprint arXiv:1706.04336*, 2017.

[280] N. V. Chawla, "Data mining for imbalanced datasets: An overview," *Data mining and knowledge discovery handbook*, pp. 875–886, 2009.

[281] S. Uddin, A. Khan, M. E. Hossain, and M. A. Moni, "Comparing different supervised machine learning algorithms for disease prediction," *BMC medical informatics and decision making*, vol. 19, no. 1, pp. 1–16, 2019.

[282] B. Filtjens, A. Nieuwboer, N. D'cruz, J. Spildooren, P. Slaets, and B. Vanrumste, "A data-driven approach for detecting gait events during turning in people with parkinson's disease and freezing of gait," *Gait & Posture*, vol. 80, pp. 130–136, 2020.

[283] R. Lamba, T. Gulati, K. A. Al-Dhlan, and A. Jain, "A systematic approach to diagnose parkinson's disease through kinematic features extracted from handwritten drawings," *Journal of Reliable Intelligent Environments*, pp. 1–10, 2021.

[284] P. Ghaderyan, A. Abbasi, and S. Saber, "A new algorithm for kinematic analysis of handwriting data; towards a reliable handwriting-based tool for early detection of alzheimer's disease," *Expert Systems with Applications*, vol. 114, pp. 428–440, 2018.

[285] L. Costa, M. F. Gago, D. Yelshyna, J. Ferreira, H. David Silva, L. Rocha, N. Sousa, and E. Bicho, "Application of machine learning in postural control kinematics for the diagnosis of alzheimer's disease," *Computational intelligence and neuroscience*, vol. 2016, 2016.

[286] S. Aich, P. M. Pradhan, J. Park, and H.-C. Kim, "A machine learning approach to distinguish parkinson's disease (pd) patient's with shuffling gait from older adults based on gait signals using 3d motion analysis," *Int. J. Eng. Technol*, vol. 7, no. 3.29, pp. 153–156, 2018.

[287] P. Drotár, J. Mekyska, I. Rektorová, L. Masarová, Z. Smékal, and M. Faundez-Zanuy, "Analysis of in-air movement in handwriting: A novel marker for parkinson's disease," *Computer methods and programs in biomedicine*, vol. 117, no. 3, pp. 405–411, 2014.

[288] J. D. L. Duque, A. J. S. Egea, T. Reeb, H. A. G. Rojas, and A. M. Gonzalez-Vargas, "Angular velocity analysis boosted by machine learning for helping in the differential diagnosis of parkinson's disease and essential tremor," *IEEE access*, vol. 8, pp. 88866–88875, 2020.

[289] I. Tien, S. D. Glaser, and M. J. Aminoff, "Characterization of gait abnormalities in parkinson's disease using a wireless inertial sensor system," in *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, pp. 3353–3356, IEEE, 2010.

[290] S. Moon, H.-J. Song, V. D. Sharma, K. E. Lyons, R. Pahwa, A. E. Akinwuntan, and H. Devos, "Classification of parkinson's disease and essential tremor based on balance and gait characteristics from wearable motion sensors via machine learning techniques: a data-driven approach," *Journal of NeuroEngineering and Rehabilitation*, vol. 17, no. 1, pp. 1–8, 2020.

[291] M. De Vos, J. Prince, T. Buchanan, J. J. FitzGerald, and C. A. Antoniades, "Discriminating progressive supranuclear palsy from parkinson's disease using wearable technology and machine learning," *Gait & posture*, vol. 77, pp. 257–263, 2020.

[292] Y. Zhang and Y. Ma, "Application of supervised machine learning algorithms in the classification of sagittal gait patterns of cerebral palsy children with spastic diplegia," *Computers in biology and medicine*, vol. 106, pp. 33–39, 2019.

[293] A. Mannini, D. Trojaniello, A. Cereatti, and A. M. Sabatini, "A machine learning framework for gait classification using inertial sensors: Application to elderly, post-stroke and huntington's disease patients," *Sensors*, vol. 16, no. 1, p. 134, 2016.

[294] E. Park, K. Lee, T. Han, and H. S. Nam, "Automatic grading of stroke symptoms for rapid assessment using optimized machine learning and 4-limb kinematics: clinical validation study," *Journal of Medical Internet Research*, vol. 22, no. 9, p. e20641, 2020.

[295] J.-s. Cho, Y.-S. Cho, S.-B. Moon, M.-J. Kim, H. D. Lee, S. Y. Lee, Y.-H. Ji, Y.-S. Park, C.-S. Han, and S.-H. Jang, "Scoliosis screening through a machine learning based gait analysis test," *International Journal of Precision Engineering and Manufacturing*, vol. 19, no. 12, pp. 1861–1872, 2018.

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

[297] W. Teufl, B. Taetz, M. Miezal, C. Dindorf, M. Fröhlich, U. Trinler, A. Hogan, and G. Bleser, "Automated detection and explainability of pathological gait patterns using a one-class support vector machine trained on inertial measurement unit based gait data," *Clinical Biomechanics*, vol. 89, p. 105452, 2021.

[298] J.-Y. Kim, G. Park, S.-A. Lee, and Y. Nam, "Analysis of machine learning-based assessment for elbow spasticity using inertial sensors," *Sensors*, vol. 20, no. 6, p. 1622, 2020.

[299] R. LeMoyne, F. Heerinckx, T. Aranca, R. De Jager, T. Zesiewicz, and H. J. Saal, "Wearable body and wireless inertial sensors for machine learning classification of gait for people with friedreich's ataxia," in *2016 IEEE 13th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, pp. 147–151, IEEE, 2016.

[300] E. Halilaj, A. Rajagopal, M. Fiterau, J. L. Hicks, T. J. Hastie, and S. L. Delp, "Machine learning in human movement biomechanics: Best practices, common pitfalls, and new opportunities," *Journal of biomechanics*, vol. 81, pp. 1–11, 2018.

[301] P. Drotár, J. Mekyska, I. Rektorová, L. Masarová, Z. Smékal, and M. Faundez-Zanuy, "Evaluation of handwriting kinematics and pressure for differential diagnosis of parkinson's disease," *Artificial intelligence in Medicine*, vol. 67, pp. 39–46, 2016.

[302] A. C. A. de Araújo, E. G. d. R. Santos, K. S. G. de Sá, V. K. T. Furtado, F. A. Santos, R. C. de Lima, L. V. Krejcová, B. L. Santos-Lobato, G. H. L. Pinto, A. d. S. Cabral, *et al.*, "Hand resting tremor assessment of healthy and patients with parkinson's disease: an exploratory machine learning study," *Frontiers in bioengineering and biotechnology*, vol. 8, p. 778, 2020.

[303] H. Jeon, W. Lee, H. Park, H. J. Lee, S. K. Kim, H. B. Kim, B. Jeon, and K. S. Park, "High-accuracy automatic classification of parkinsonian tremor severity using machine learning method," *Physiological measurement*, vol. 38, no. 11, p. 1980, 2017.

[304] C. Caramia, D. Torricelli, M. Schmid, A. Munoz-Gonzalez, J. Gonzalez-Vargas, F. Grandas, and J. L. Pons, "Imu-based classification of parkinson's disease from gait: A sensitivity analysis

on sensor location and feature selection," *IEEE journal of biomedical and health informatics*, vol. 22, no. 6, pp. 1765–1774, 2018.

[305] D. J. Park, J. W. Lee, M. J. Lee, S. J. Ahn, J. Kim, G. L. Kim, Y. J. Ra, Y. N. Cho, and W. B. Jeong, "Evaluation for parkinsonian bradykinesia by deep learning modeling of kinematic parameters," *Journal of Neural Transmission*, vol. 128, no. 2, pp. 181–189, 2021.

[306] B. P. Printy, L. M. Renken, J. P. Herrmann, I. Lee, B. Johnson, E. Knight, G. Varga, and D. Whitmer, "Smartphone application for classification of motor impairment severity in parkinson's disease," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 2686–2689, IEEE, 2014.

[307] F. Demrozi, G. Pravadelli, A. Bihorac, and P. Rashidi, "Human activity recognition using inertial, physiological and environmental sensors: A comprehensive survey," *IEEE Access*, vol. 8, pp. 210816–210836, 2020.

[308] A. A. Badawi, A. Al-Kabbany, and H. Shaban, "Multimodal human activity recognition from wearable inertial sensors using machine learning," in *2018 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES)*, pp. 402–407, IEEE, 2018.

[309] T. Zebin, P. J. Scully, and K. B. Ozanyan, "Evaluation of supervised classification algorithms for human activity recognition with inertial sensors," in *2017 IEEE SENSORS*, pp. 1–3, IEEE, 2017.

[310] M. M. Hassan, M. Z. Uddin, A. Mohamed, and A. Almogren, "A robust human activity recognition system using smartphone sensors and deep learning," *Future Generation Computer Systems*, vol. 81, pp. 307–313, 2018.

[311] K. Nurhanim, I. Elamvazuthi, L. Izhar, and T. Ganesan, "Classification of human activity based on smartphone inertial sensor using support vector machine," in *2017 ieee 3rd international symposium in robotics and manufacturing automation (roma)*, pp. 1–5, IEEE, 2017.

[312] L. Xie, J. Tian, G. Ding, and Q. Zhao, "Human activity recognition method based on inertial sensor and barometer," in *2018 IEEE International Symposium on Inertial Sensors and Systems (INERTIAL)*, pp. 1–4, IEEE, 2018.

[313] T. Zebin, P. J. Scully, and K. B. Ozanyan, "Human activity recognition with inertial sensors using a deep learning approach," in *2016 IEEE SENSORS*, pp. 1–3, IEEE, 2016.

[314] A. E. Minarno, W. A. Kusuma, and H. Wibowo, "Performance comparisson activity recognition using logistic regression and support vector machine," in *2020 3rd International Conference on Intelligent Autonomous Systems (ICoIAS)*, pp. 19–24, IEEE, 2020.

[315] S. Scheurer, S. Tedesco, K. N. Brown, and B. O'Flynn, "Human activity recognition for emergency first responders via body-worn inertial sensors," in *2017 IEEE 14th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, pp. 5–8, IEEE, 2017.

[316] Y.-L. Hsu, S.-C. Yang, H.-C. Chang, and H.-C. Lai, "Human daily and sport activity recognition using a wearable inertial sensor network," *IEEE Access*, vol. 6, pp. 31715–31728, 2018.

[317] X. Yu, H. Qiu, and S. Xiong, "A novel hybrid deep neural network to predict pre-impact fall for older people based on wearable inertial sensors," *Frontiers in bioengineering and biotechnology*, vol. 8, p. 63, 2020.

[318] C. Tunca, G. Salur, and C. Ersoy, "Deep learning for fall risk assessment with inertial sensors: Utilizing domain knowledge in spatio-temporal gait parameters," *IEEE journal of biomedical and health informatics*, vol. 24, no. 7, pp. 1994–2005, 2019.

[319] A. T. Özdemir and B. Barshan, "Detecting falls with wearable sensors using machine learning techniques," *Sensors*, vol. 14, no. 6, pp. 10691–10708, 2014.

[320] A. Delval, N. Betrouni, C. Tard, D. Devos, K. Dujardin, L. Defebvre, J. Labidi, and C. Moreau, "Do kinematic gait parameters help to discriminate between fallers and non-fallers with parkinson's disease?," *Clinical neurophysiology*, vol. 132, no. 2, pp. 536–541, 2021.

[321] N. Nahar, M. S. Hossain, and K. Andersson, "A machine learning based fall detection for elderly people with neurodegenerative disorders," in *International Conference on Brain Informatics*, pp. 194–203, Springer, 2020.

[322] S. Gillain, M. Boutaayamou, C. Schwartz, O. Brüls, O. Bruyère, J.-L. Croisier, E. Salmon, J.-Y. Reginster, G. Garraux, and J. Petermans, "Using supervised learning machine algorithm to identify future fallers based on gait patterns: a two-year longitudinal study," *Experimental gerontology*, vol. 127, p. 110730, 2019.

[323] M. Nahiduzzaman, M. Tasnim, N. T. Newaz, M. S. Kaiser, and M. Mahmud, "Machine learning based early fall detection for elderly people with neurological disorder using multimodal data fusion," in *International Conference on Brain Informatics*, pp. 204–214, Springer, 2020.

[324] E. Ribeiro, L. Bentes, A. Cruz, G. Leitão, R. Barreto, V. Silva, T. Primo, and F. Koch, "On the use of inertial sensors and machine learning for automatic recognition of fainting and epileptic seizure," in *2016 IEEE 18th International Conference on e-Health Networking, Applications and Services (Healthcom)*, pp. 1–6, IEEE, 2016.

[325] H.-Y. Lau, K.-Y. Tong, and H. Zhu, "Support vector machine for classification of walking conditions using miniature kinematic sensors," *Medical & biological engineering & computing*, vol. 46, no. 6, pp. 563–573, 2008.

[326] N. Ghazali, N. Shahar, N. Rahmad, N. J. Sufri, M. As' ari, and H. M. Latif, "Common sport activity recognition using inertial sensor," in *2018 IEEE 14th International Colloquium on Signal Processing & Its Applications (CSPA)*, pp. 67–71, IEEE, 2018.

[327] D. F. Whelan, M. A. O'Reilly, T. E. Ward, E. Delahunt, and B. Caulfield, "Technology in rehabilitation: Comparing personalised and global classification methodologies in evaluating the squat exercise with wearable imus," *Methods of information in medicine*, vol. 56, no. 05, pp. 361–369, 2017.

[328] O. Giggins, K. T. Sweeney, and B. Caulfield, "The use of inertial sensors for the classification of rehabilitation exercises," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 2965–2968, IEEE, 2014.

[329] P.-C. Chen, C.-N. Huang, I.-C. Chen, and C.-T. Chan, "A rehabilitation exercise assessment system based on wearable sensors for knee osteoarthritis," in *International Conference on Smart Homes and Health Telematics*, pp. 267–272, Springer, 2013.

[330] M. A. O'Reilly, D. F. Whelan, T. E. Ward, E. Delahunt, and B. M. Caulfield, "Classification of deadlift biomechanics with wearable inertial measurement units," *Journal of biomechanics*, vol. 58, pp. 155–161, 2017.

[331] P. E. Taylor, G. J. Almeida, T. Kanade, and J. K. Hodgins, "Classifying human motion quality for knee osteoarthritis using accelerometers," in *2010 Annual international conference of the IEEE engineering in medicine and biology*, pp. 339–343, IEEE, 2010.

[332] Y. Zheng, "Human activity recognition based on the hierarchical feature selection and classification framework," *Journal of Electrical and Computer Engineering*, vol. 2015, 2015.

[333] B. Barshan and M. C. Yüksek, "Recognizing daily and sports activities in two open source machine learning environments using body-worn sensor units," *The Computer Journal*, vol. 57, no. 11, pp. 1649–1667, 2014.

[334] G. De Leonardis, S. Rosati, G. Balestra, V. Agostini, E. Panero, L. Gastaldi, and M. Knaflitz, "Human activity recognition by wearable sensors: Comparison of different classifiers for real-time applications," in *2018 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, pp. 1–6, IEEE, 2018.

[335] Z. Feng, L. Mo, and M. Li, "A random forest-based ensemble method for activity recognition," in *2015 37th annual international conference of the ieee engineering in medicine and biology society (embc)*, pp. 5074–5077, IEEE, 2015.

[336] R. de Pinho André, A. Raposo, and H. Fuks, "Using foot and knee movement and posture information to mitigate the probability of injuries in functional training," in *International Conference on Human-Computer Interaction*, pp. 153–169, Springer, 2019.

[337] J. Zhang, T. E. Lockhart, and R. Soangra, "Classifying lower extremity muscle fatigue during walking using machine learning and inertial sensors," *Annals of biomedical engineering*, vol. 42, no. 3, pp. 600–612, 2014.

[338] M. Guaitolini, L. Truppa, A. M. Sabatini, A. Mannini, and C. Castagna, "Sport-induced fatigue detection in gait parameters using inertial sensors and support vector machines," in *2020 8th IEEE RAS/EMBS International Conference for Biomedical Robotics and Biomechatronics (BioRob)*, pp. 170–174, IEEE, 2020.

[339] T. Op De Beéck, W. Meert, K. Schütte, B. Vanwanseele, and J. Davis, "Fatigue prediction in outdoor runners via machine learning and sensor fusion," in *Proceedings of the 24th acm sigkdd international conference on knowledge discovery & data mining*, pp. 606–615, 2018.

[340] L. Marotta, J. H. Buurke, B.-J. F. van Beijnum, and J. Reenalda, "Towards machine learning-based detection of running-induced fatigue in real-world scenarios: Evaluation of imu sensor configurations to reduce intrusiveness," *Sensors*, vol. 21, no. 10, p. 3451, 2021.

[341] C. Buckley, M. A. O'Reilly, D. Whelan, A. V. Farrell, L. Clark, V. Longo, M. Gilchrist, and B. Caulfield, "Binary classification of running fatigue using a single inertial measurement unit," in *2017 IEEE 14th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, pp. 197–201, IEEE, 2017.

[342] Q. Liu, S. Mo, V. C. Cheung, B. M. Cheung, S. Wang, P. P. Chan, A. Malhotra, R. T. Cheung, and R. H. Chan, "Classification of runners' performance levels with concurrent prediction of biomechanical parameters using data from inertial measurement units," *Journal of Biomechanics*, vol. 112, p. 110072, 2020.

[343] C. Strohrmann, M. Rossi, B. Arnrich, and G. Troster, "A data-driven approach to kinematic analysis in running using wearable technology," in *2012 Ninth International Conference on Wearable and Implantable Body Sensor Networks*, pp. 118–123, IEEE, 2012.

[344] L. C. Benson, C. A. Clermont, S. T. Osis, D. Kobsar, and R. Ferber, "Classifying running speed conditions using a single wearable sensor: Optimal segmentation and feature extraction methods," *Journal of biomechanics*, vol. 71, pp. 94–99, 2018.

[345] P. Dixon, K. Schütte, B. Vanwanseele, J. Jacobs, J. Dennerlein, J. Schiffman, P. Fournier, and B. Hu, "Machine learning algorithms can classify outdoor terrain types during running using accelerometry data," *Gait & posture*, vol. 74, pp. 176–181, 2019.

[346] R. Argent, S. Drummond, A. Remus, M. O'Reilly, and B. Caulfield, "Evaluating the use of machine learning in the assessment of joint angle using a single inertial sensor," *Journal of rehabilitation and assistive technologies engineering*, vol. 6, p. 2055668319868544, 2019.

[347] H. Lim, B. Kim, and S. Park, "Prediction of lower limb kinetics and kinematics during walking by a single imu on the lower back using machine learning," *Sensors*, vol. 20, no. 1, p. 130, 2020.

[348] M. Mundt, W. Thomsen, T. Witter, A. Koeppe, S. David, F. Bamer, W. Potthast, and B. Markert, "Prediction of lower limb joint angles and moments during gait using artificial neural networks," *Medical & biological engineering & computing*, vol. 58, no. 1, pp. 211–225, 2020.

[349] C. R. Chaaban, N. T. Berry, C. Armitano-Lago, A. W. Kiefer, M. J. Mazzoleni, and D. A. Padua, "Combining inertial sensors and machine learning to predict vgrf and knee biomechanics during a double limb jump landing task," *Sensors*, vol. 21, no. 13, p. 4383, 2021.

[350] F. J. Wouda, M. Giuberti, G. Bellusci, E. Maartens, J. Reenalda, B.-J. F. Van Beijnum, and P. H. Veltink, "Estimation of vertical ground reaction forces and sagittal knee kinematics during running using three inertial sensors," *Frontiers in physiology*, vol. 9, p. 218, 2018.

[351] X. Jiang, M. Gholami, M. Khoshnam, J. J. Eng, and C. Menon, "Estimation of ankle joint power during walking using two inertial sensors," *Sensors*, vol. 19, no. 12, p. 2796, 2019.

[352] R. Derie, P. Robberechts, P. Van den Berghe, J. Gerlo, D. De Clercq, V. Segers, and J. Davis, "Tibial acceleration-based prediction of maximal vertical loading rate during overground running: a machine learning approach," *Frontiers in bioengineering and biotechnology*, vol. 8, p. 33, 2020.

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

[354] M. Gholami, C. Napier, and C. Menon, "Estimating lower extremity running gait kinematics with a single accelerometer: A deep learning approach," *Sensors*, vol. 20, no. 10, p. 2939, 2020.

[355] D. H. K. Chow, L. Tremblay, C. Y. Lam, A. W. Y. Yeung, W. H. W. Cheng, and P. T. W. Tse, "Comparison between accelerometer and gyroscope in predicting level-ground running kinematics by treadmill running kinematics using a single wearable sensor," *Sensors*, vol. 21, no. 14, p. 4633, 2021.

[356] H. Zhang, Y. Guo, and D. Zanotto, "Accurate ambulatory gait analysis in walking and running using machine learning models," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 1, pp. 191–202, 2019.

[357] A. De Brabandere, J. Emmerzaal, A. Timmermans, I. Jonkers, B. Vanwanseele, and J. Davis, "A machine learning approach to estimate hip and knee joint loading using a mobile phone-embedded imu," *Frontiers in bioengineering and biotechnology*, vol. 8, p. 320, 2020.

[358] B. J. Stetter, F. C. Krafft, S. Ringhof, T. Stein, and S. Sell, "A machine learning and wearable sensor based approach to estimate external knee flexion and adduction moments during various locomotion tasks," *Frontiers in bioengineering and biotechnology*, vol. 8, p. 9, 2020.

[359] A. I. Cuesta-Vargas, A. Galán-Mercant, and J. M. Williams, "The use of inertial sensors system for human motion analysis," *Physical Therapy Reviews*, vol. 15, no. 6, pp. 462–473, 2010.

[360] B. T. Saragiotto, T. P. Yamato, and A. D. Lopes, "What do recreational runners think about risk factors for running injuries? a descriptive study of their beliefs and opinions," *journal of orthopaedic & sports physical therapy*, vol. 44, no. 10, pp. 733–738, 2014.

[361] B. M. Marshall, A. D. Franklyn-Miller, E. A. King, K. A. Moran, S. C. Strike, and É. C. Falvey, "Biomechanical factors associated with time to complete a change of direction cutting maneuver," *The Journal of Strength & Conditioning Research*, vol. 28, no. 10, pp. 2845–2851, 2014.

[362] S. Mo and D. H. K. Chow, "Differences in lower-limb coordination and coordination variability between novice and experienced runners during a prolonged treadmill run at anaerobic threshold speed," *Journal of Sports Sciences*, vol. 37, no. 9, pp. 1021–1028, 2019.

[363] R. M. Koldenhoven and J. Hertel, "Validation of a wearable sensor for measuring running biomechanics," *Digital biomarkers*, vol. 2, no. 2, pp. 74–78, 2018.

[364] N. Chambon, N. Delattre, N. Guéguen, E. Berton, and G. Rao, "Shoe drop has opposite influence on running pattern when running overground or on a treadmill," *European journal of applied physiology*, vol. 115, no. 5, pp. 911–918, 2015.

[365] D. Oriwol, T. L. Milani, and C. Maiwald, "Methodological issues associated with the mean value of repeated laboratory running measurements," *Footwear Science*, vol. 4, no. 3, pp. 183–190, 2012.

[366] M. B. Del Rosario, N. H. Lovell, and S. J. Redmond, "Quaternion-based complementary filter for attitude determination of a smartphone," *IEEE Sensors Journal*, vol. 16, no. 15, pp. 6008–6017, 2016.

[367] T. J. Hullfish, F. Qu, B. D. Stoeckl, P. M. Gebhard, R. L. Mauck, and J. R. Baxter, "Measuring clinically relevant knee motion with a self-calibrated wearable sensor," *Journal of biomechanics*, vol. 89, pp. 105–109, 2019.

[368] N. Yadav and C. Bleakley, "Accurate orientation estimation using ahrs under conditions of magnetic distortion," *Sensors*, vol. 14, no. 11, pp. 20008–20024, 2014.

[369] B. Zhou, J. Yang, and Q. Li, "Smartphone-based activity recognition for indoor localization using a convolutional neural network," *Sensors*, vol. 19, no. 3, p. 621, 2019.

[370] E. Dorschky, M. Nitschke, C. F. Martindale, A. J. van den Bogert, A. D. Koelewijn, and B. M. Eskofier, "Cnn-based estimation of sagittal plane walking and running biomechanics from measured and simulated inertial sensor data," *Frontiers in bioengineering and biotechnology*, vol. 8, p. 604, 2020.

[371] M. Munoz-Organero, "Outlier detection in wearable sensor data for human activity recognition (har) based on drnns," *IEEE Access*, vol. 7, pp. 74422–74436, 2019.

[372] S. Madgwick *et al.*, "An efficient orientation filter for inertial and inertial/magnetic sensor arrays," *Report x-io and University of Bristol (UK)*, vol. 25, pp. 113–118, 2010.

[373] X. Xiao and S. Zarar, "Packet loss concealment with recurrent neural networks for wireless inertial pose tracking," in *2018 IEEE 15th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, pp. 25–29, IEEE, 2018.

[374] T. Hossain, M. Ahad, A. Rahman, and S. Inoue, "A method for sensor-based activity recognition in missing data scenario," *Sensors*, vol. 20, no. 14, p. 3811, 2020.

[375] T. Feng and S. Narayanan, "Imputing missing data in large-scale multivariate biomedical wearable recordings using bidirectional recurrent neural networks with temporal activation regularization," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 2529–2534, IEEE, 2019.

[376] S. Lin, X. Wu, G. Martinez, and N. V. Chawla, "Filling missing values on wearable-sensory time series data," in *Proceedings of the 2020 SIAM International Conference on Data Mining*, pp. 46–54, SIAM, 2020.

[377] I. Azimi, T. Pahikkala, A. M. Rahmani, H. Niela-Vilén, A. Axelin, and P. Liljeberg, "Missing data resilient decision-making for healthcare iot through personalization: A case study on maternal health," *Future Generation Computer Systems*, vol. 96, pp. 297–308, 2019.

[378] G. Mariethoz, N. Linde, D. Jougnot, and H. Rezaee, "Feature-preserving interpolation and filtering of environmental time series," *Environmental Modelling & Software*, vol. 72, pp. 71–76, 2015.

[379] P. Körner, R. Kronenberg, S. Genzel, and C. Bernhofer, "Introducing gradient boosting as a universal gap filling tool for meteorological time series," *Meteorologische Zeitschrift*, vol. 27, no. 5, pp. 369–376, 2018.

[380] B.-S. Lin, Y.-S. Lin, I.-J. Lee, and B.-S. Lin, "A real-time missing data recovery method using recurrent neural network for multiple transmissions," in *International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, pp. 99–107, Springer, 2018.

[381] Z. Che, S. Purushotham, K. Cho, D. Sontag, and Y. Liu, "Recurrent neural networks for multivariate time series with missing values," *Scientific reports*, vol. 8, no. 1, pp. 1–12, 2018.

[382] K. J. Kim, V. Agrawal, I. Gaunaurd, R. S. Gailey, and C. L. Bennett, "Missing sample recovery for wireless inertial sensor-based human movement acquisition," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 24, no. 11, pp. 1191–1198, 2016.

[383] L. De Vito, G. Mazzilli, M. Riccio, and C. Sementa, "Improving the orientation estimation in a packet loss-affected wireless sensor network for tracking human motion," in *Proceedings of 21st IMEKO TC4 International Symposium on Understanding the World through Electrical and Electronic Measurement*, pp. 184–189, 2016.

[384] D. Morelli, A. Rossi, M. Cairo, and D. A. Clifton, "Analysis of the impact of interpolation methods of missing rr-intervals caused by motion artifacts on hrv features estimations," *Sensors*, vol. 19, no. 14, p. 3163, 2019.

[385] J. Van de Molengraft, S. Nimmala, B. Mariani, K. Aminian, and J. Penders, "Wireless 6d inertial measurement platform for ambulatory gait monitoring," in *Proc. of the 6th Int. workshop on Wearable, Micro and Nanosystems for Personalised Health*, pp. 63–64, 2009.

[386] J. Reenalda, M. A. Zandbergen, J. H. Harbers, M. R. Paquette, and C. E. Milner, "Detection of foot contact in treadmill running with inertial and optical measurement systems," *Journal of Biomechanics*, vol. 121, p. 110419, 2021.

[387] F. J. Wouda, S. L. Jaspar, J. Harlaar, B.-J. F. van Beijnum, and P. H. Veltink, "Foot progression angle estimation using a single foot-worn inertial sensor," *Journal of neuroengineering and rehabilitation*, vol. 18, no. 1, pp. 1–10, 2021.

[388] F. Höflinger, J. Müller, R. Zhang, L. M. Reindl, and W. Burgard, "A wireless micro inertial measurement unit (imu)," *IEEE Transactions on instrumentation and measurement*, vol. 62, no. 9, pp. 2583–2595, 2013.

[389] C. Neville, C. Ludlow, and B. Rieger, "Measuring postural stability with an inertial sensor: validity and sensitivity," *Medical devices (Auckland, NZ)*, vol. 8, p. 447, 2015.

[390] C. M. Kanzler, J. Barth, A. Rampp, H. Schlarb, F. Rott, J. Klucken, and B. M. Eskofier, "Inertial sensor based and shoe size independent gait analysis including heel and toe clearance estimation," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 5424–5427, IEEE, 2015.

[391] C. Pappas, S. M. Papalexiou, and D. Koutsoyiannis, "A quick gap filling of missing hydrometeorological data," *Journal of Geophysical Research: Atmospheres*, vol. 119, no. 15, pp. 9290–9300, 2014.

[392] L. Pashova, P. Koprinkova-Hristova, and S. Popova, "Gap filling of daily sea levels by artificial neural networks," *TransNav: International Journal on Marine Navigation and Safety of Sea Transportation*, vol. 7, no. 2, 2013.

[393] H. Junninen, H. Niska, K. Tuppurainen, J. Ruuskanen, and M. Kolehmainen, "Methods for imputation of missing values in air quality data sets," *Atmospheric Environment*, vol. 38, no. 18, pp. 2895–2907, 2004.

[394] B. Staub, A. Hasler, J. Noetzli, and R. Delaloye, "Gap-filling algorithm for ground surface temperature data measured in permafrost and periglacial environments," *Permafrost and Periglacial Processes*, vol. 28, no. 1, pp. 275–285, 2017.

[395] A. Janssen, R. Veldhuis, and L. Vries, "Adaptive interpolation of discrete-time signals that can be modeled as autoregressive processes," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 34, no. 2, pp. 317–330, 1986.

[396] J. C. Baltazar and D. E. Claridge, "Restoration of short periods of missing energy use and weather data using cubic spline and fourier series approaches: qualitative analysis," 2002.

[397] F. Avanzi, Z. Zheng, A. Coogan, R. Rice, R. Akella, and M. H. Conklin, "Gap-filling snow-depth time-series with kalman filtering-smoothing and expectation maximization: Proof of concept

using spatially dense wireless-sensor-network data," *Cold Regions Science and Technology*, vol. 175, p. 103066, 2020.

[398] J. P. Musial, M. M. Verstraete, and N. Gobron, "Comparing the effectiveness of recent algorithms to fill and smooth incomplete and noisy time series," *Atmospheric chemistry and physics*, vol. 11, no. 15, pp. 7905–7923, 2011.

[399] S. Khelifa, B. Gourine, H. Taibi, and H. Dekkiche, "Filling gaps in time series of space-geodetic positioning," *Arabian Journal of Geosciences*, vol. 11, no. 12, pp. 1–7, 2018.

[400] A. Verger, F. Baret, M. Weiss, S. Kandasamy, and E. Vermote, "The cacao method for smoothing, gap filling, and characterizing seasonal anomalies in satellite time series," *IEEE transactions on Geoscience and Remote Sensing*, vol. 51, no. 4, pp. 1963–1972, 2013.

[401] D. Rasch and V. Guiard, "The robustness of parametric statistical methods," *Psychology Science*, vol. 46, pp. 175–208, 2004.

[402] M. J. Blanca Mena, R. Alarcón Postigo, J. Arnau Gras, R. Bono Cabré, and R. Bendayan, "Non-normal data: Is anova still a valid option?," *Psicothema, 2017, vol. 29, num. 4, p. 552-557*, 2017.

[403] M. Lepot, J.-B. Aubin, and F. H. Clemens, "Interpolation in time series: An introductive overview of existing methods, their performance criteria and uncertainty assessment," *Water*, vol. 9, no. 10, p. 796, 2017.

[404] T.-B. Ottosen and P. Kumar, "Outlier detection and gap filling methodologies for low-cost air quality measurements," *Environmental Science: Processes & Impacts*, vol. 21, no. 4, pp. 701–713, 2019.

[405] M. Nazarahari and H. Rouhani, "Adaptive gain regulation of sensor fusion algorithms for orientation estimation with magnetic and inertial measurement units," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–13, 2020.

[406] B. T. Polyak, "Some methods of speeding up the convergence of iteration methods," *Ussr computational mathematics and mathematical physics*, vol. 4, no. 5, pp. 1–17, 1964.

[407] I. Sutskever, J. Martens, G. Dahl, and G. Hinton, "On the importance of initialization and momentum in deep learning," in *International conference on machine learning*, pp. 1139–1147, PMLR, 2013.

[408] A. Cutkosky and H. Mehta, "Momentum improves normalized sgd," in *International conference on machine learning*, pp. 2260–2268, PMLR, 2020.

[409] H. Wang, J. Zhao, J. Li, L. Tian, P. Tu, T. Cao, Y. An, K. Wang, and S. Li, "Wearable sensor-based human activity recognition using hybrid deep learning techniques," *Security and communication Networks*, vol. 2020, 2020.

[410] C. E. Rothschild, "Primitive running: a survey analysis of runners' interest, participation, and implementation," *The Journal of Strength & Conditioning Research*, vol. 26, no. 8, pp. 2021–2026, 2012.

[411] D. E. Warburton, C. W. Nicol, and S. S. Bredin, "Health benefits of physical activity: the evidence," *Cmaj*, vol. 174, no. 6, pp. 801–809, 2006.

[412] I. Paul, M. B. Munro, P. Abernethy, S. Simon, E. Radin, and R. Rose, "Musculo-skeletal shock absorption: relative contribution of bone and soft tissues at various frequencies," *Journal of biomechanics*, vol. 11, no. 5, pp. 237–239, 1978.

[413] Q. Louw and K. Grimmer, "Biomechanical factors associated with the risk of knee injury when landing from a jump," *South African Journal of Sports Medicine*, vol. 18, no. 1, pp. 18–23, 2006.

[414] D. E. Lieberman, "What we can learn about running from barefoot running: an evolutionary medical perspective," *Exercise and sport sciences reviews*, vol. 40, no. 2, pp. 63–72, 2012.

[415] D. Karen, D. S. Green, G. M. Jensen, and E. Savinar, "A comparison of muscular tightness in runners and nonrunners and the relation of muscular tightness to low back pain in runners," *Journal of Orthopaedic & Sports Physical Therapy*, vol. 6, no. 6, pp. 315–323, 1985.

[416] R. Bahr, "Why screening tests to predict injury do not work—and probably never will...: a critical review," *British journal of sports medicine*, vol. 50, no. 13, pp. 776–780, 2016.

[417] J. D. Ruddy, S. J. Cormack, R. Whiteley, M. D. Williams, R. G. Timmins, and D. A. Opar, "Modeling the risk of team sport injuries: a narrative review of different statistical approaches," *Frontiers in physiology*, vol. 10, p. 829, 2019.

[418] T. P. Yamato, B. T. Saragiotto, L. C. Hespanhol, S. S. Yeung, and A. D. Lopes, "Descriptors used to define running-related musculoskeletal injury: a systematic review," *journal of orthopaedic & sports physical therapy*, vol. 45, no. 5, pp. 366–374, 2015.

[419] S. Varma and R. Simon, "Bias in error estimation when using cross-validation for model selection," *BMC bioinformatics*, vol. 7, no. 1, pp. 1–8, 2006.

[420] S. Raschka, "Model evaluation, model selection, and algorithm selection in machine learning," *arXiv preprint arXiv:1811.12808*, 2018.

[421] G. M. Sullivan and R. Feinn, "Using effect size—or why the p value is not enough," *Journal of graduate medical education*, vol. 4, no. 3, pp. 279–282, 2012.

[422] J. Cohen, *Statistical power analysis for the behavioral sciences.* Academic press, 2013.

[423] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, *et al.*, "Scikit-learn: Machine learning in python," *the Journal of machine Learning research*, vol. 12, pp. 2825–2830, 2011.

[424] M. A. O'Reilly, W. Johnston, C. Buckley, D. Whelan, and B. Caulfield, "The influence of feature selection methods on exercise classification with inertial measurement units," in *2017 IEEE 14th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, pp. 193–196, IEEE, 2017.

[425] R. Wazirali, "An improved intrusion detection system based on knn hyperparameter tuning and cross-validation," *Arabian Journal for Science and Engineering*, vol. 45, no. 12, pp. 10859–10873, 2020.

[426] A. Davoudi, A. A. Wanigatunga, M. Kheirkhahan, D. B. Corbett, T. Mendoza, M. Battula, S. Ranka, R. B. Fillingim, T. M. Manini, P. Rashidi, *et al.*, "Accuracy of samsung gear s smartwatch for activity recognition: Validation study," *JMIR mHealth and uHealth*, vol. 7, no. 2, p. e11270, 2019.

[427] I. H. Witten, E. Frank, M. A. Hall, C. Pal, and M. DATA, "Practical machine learning tools and techniques," in *DATA MINING*, vol. 2, p. 4, 2005.

[428] L. Sanhudo, D. Calvetti, J. P. Martins, N. M. Ramos, P. Mêda, M. C. Gonçalves, and H. Sousa, "Activity classification using accelerometers and machine learning for complex construction worker activities," *Journal of Building Engineering*, vol. 35, p. 102001, 2021.

[429] D. Baumann and K. Baumann, "Reliable estimation of prediction errors for qsar models under model uncertainty using double cross-validation," *Journal of cheminformatics*, vol. 6, no. 1, pp. 1–19, 2014.

[430] M. Tahkola and Z. Guangrong, "Atsc-nex: Automated time series classification with sequential model-based optimization and nested cross-validation," *IEEE Access*, vol. 10, pp. 39299–39312, 2022.

[431] S. Parvandeh, H.-W. Yeh, M. P. Paulus, and B. A. McKinney, "Consensus features nested cross-validation," *Bioinformatics*, vol. 36, no. 10, pp. 3093–3098, 2020.

[432] V. Hernandez, T. Suzuki, and G. Venture, "Convolutional and recurrent neural network for human activity recognition: Application on american sign language," *PloS one*, vol. 15, no. 2, p. e0228869, 2020.

[433] T. Hastie, R. Tibshirani, J. H. Friedman, and J. H. Friedman, *The elements of statistical learning: data mining, inference, and prediction*, vol. 2. Springer, 2009.

[434] D. K. Lee, "Alternatives to p value: confidence interval and effect size," *Korean journal of anesthesiology*, vol. 69, no. 6, p. 555, 2016.

[435] D. Y. Wen, J. C. Puffer, and T. P. Schmalzried, "Lower extremity alignment and risk of overuse injuries in runners.," *Medicine and science in sports and exercise*, vol. 29, no. 10, pp. 1291–1298, 1997.

[436] T. E. Hewett, G. D. Myer, K. R. Ford, R. S. Heidt Jr, A. J. Colosimo, S. G. McLean, A. J. Van den Bogert, M. V. Paterno, and P. Succop, "Biomechanical measures of neuromuscular control and valgus loading of the knee predict anterior cruciate ligament injury risk in female athletes: a prospective study," *The American journal of sports medicine*, vol. 33, no. 4, pp. 492–501, 2005.

[437] T. Krosshaug, K. Steffen, E. Kristianslund, A. Nilstad, K.-M. Mok, G. Myklebust, T. E. Andersen, I. Holme, L. Engebretsen, and R. Bahr, "The vertical drop jump is a poor screening

test for acl injuries in female elite soccer and handball players: a prospective cohort study of 710 athletes," *The American journal of sports medicine*, vol. 44, no. 4, pp. 874–883, 2016.

[438] F. Alenezi, L. Herrington, P. Jones, and R. Jones, "How reliable are lower limb biomechanical variables during running and cutting tasks," *Journal of Electromyography and Kinesiology*, vol. 30, pp. 137–142, 2016.

[439] T. Hein, P. Janssen, U. Wagner-Fritz, G. Haupt, and S. Grau, "Prospective analysis of intrinsic and extrinsic risk factors on the development of a chilles tendon pain in runners," *Scandinavian Journal of Medicine & Science in Sports*, vol. 24, no. 3, pp. e201–e212, 2014.

[440] A. Franklyn-Miller, C. Richter, E. King, S. Gore, K. Moran, S. Strike, and E. Falvey, "Athletic groin pain (part 2): a prospective cohort study on the biomechanical evaluation of change of direction identifies three clusters of movement patterns," *British journal of sports medicine*, vol. 51, no. 5, pp. 460–468, 2017.

[441] A. Phinyomark, S. Osis, B. A. Hettinga, and R. Ferber, "Kinematic gait patterns in healthy runners: A hierarchical cluster analysis," *Journal of biomechanics*, vol. 48, no. 14, pp. 3897–3904, 2015.

[442] C. E. Milner, J. L. Hawkins, and K. G. Aubol, "Tibial acceleration during running is higher in field testing than indoor testing.," *Medicine and Science in Sports and Exercise*, vol. 52, no. 6, pp. 1361–1366, 2020.

[443] K. Thorborg, K. K. Krommes, E. Esteve, M. B. Clausen, E. M. Bartels, and M. S. Rathleff, "Effect of specific exercise-based football injury prevention programmes on the overall injury rate in football: a systematic review and meta-analysis of the fifa 11 and 11+ programmes," *British journal of sports medicine*, vol. 51, no. 7, pp. 562–571, 2017.

[444] B. E. Schlingermann, C. A. Lodge, C. Gissane, and P. M. Rankin, "Effects of the gaelic athletic association 15 on lower extremity injury incidence and neuromuscular functional outcomes in collegiate gaelic games," *The Journal of Strength & Conditioning Research*, vol. 32, no. 7, pp. 1993–2001, 2018.

Chapter 11

Appendix

11.1   Appendix A: Data-Loss Preliminary Experiment

Aim

To get a better insight into the magnitude of sensor data loss, a preliminary experiment was designed to determine the average number of consecutive samples lost, and to determine whether the number of instances of data loss changed over time.

Methodology

The data loss which occurred during the experiment in Section 7.3 was examined. Ten participants were selected at random from the dataset, and a program was designed to analyse the magnitude of data loss in each segment (tibia, thigh, sacrum) of all 10 participants.

For each segment, the loss of data was analysed over 5 minute windows throughout the recorded data (e.g. 0-5 minutes into the recording, 5-10 minutes, ..., 55-60 minutes) and determined the number of times (in each 5 minute window) that data loss occurred. This was to examine whether data loss was occurring mainly in certain times during the recording i.e. mainly at the start once the sensors were synchronised, or whether this was uniform throughout. The average maximum gap size across participants for each 5 minute window was also investigated in order to understand the average magnitudes of the highest data losses the sensors were experiencing. Finally, the number of occurrences of fixed gap sizes data losses were examined e.g. how many times 0-4 consecutive samples were missing, 5-9 consecutive samples, ..., 195 to 199 consecutive samples, and finally >200 consecutive samples. This was to get a sense of how common the higher ranges of gap sizes were occurring, as this averaging across participants removed the effect of one-off outliers in the higher gap size ranges. Two-hundred samples were chosen as an upper limit as the sampling rate was approximately 400 samples, so this would represent approximately half of the new samples per second to be lost, which

1

would contain important information.
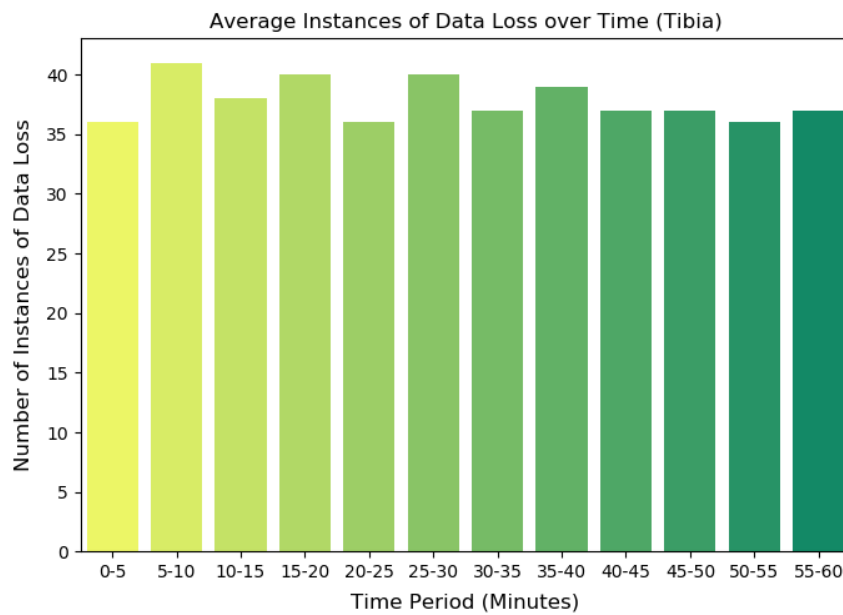
Figure 11.1: The average number of instances of data loss which occurred in each 5 minute period across participant's tibia data.
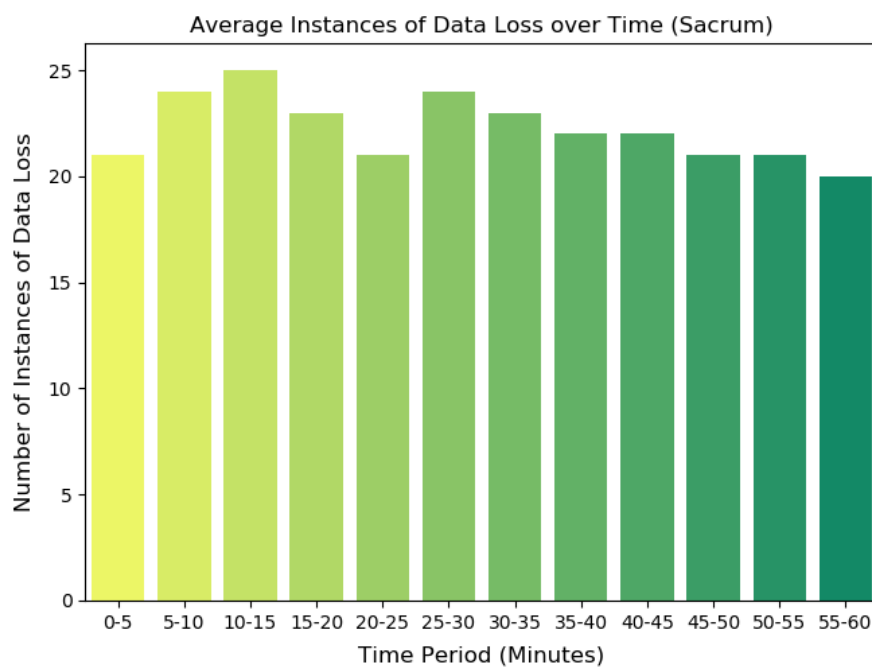
Figure 11.2: The average number of instances of data loss which occurred in each 5 minute period across participant's sacrum data.
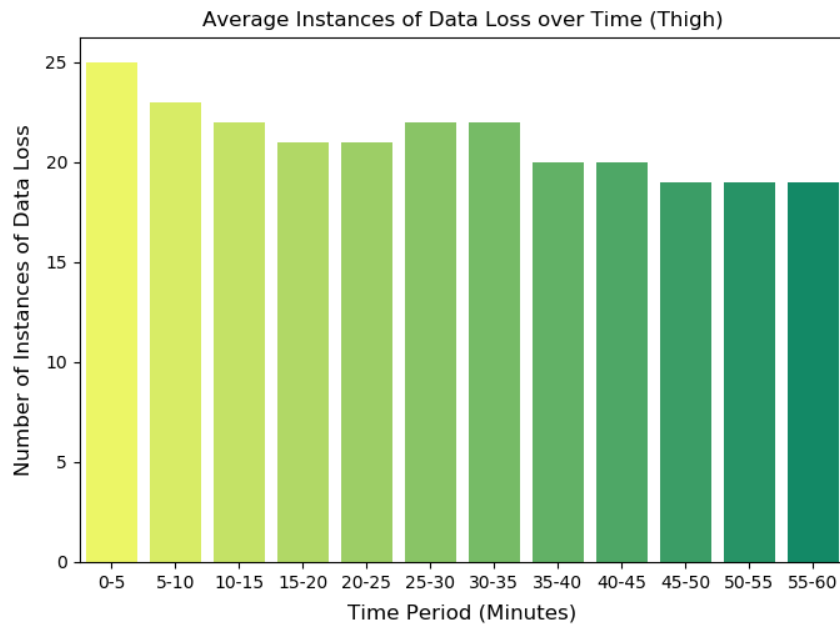
Figure 11.3: The average number of instances of data loss which occurred in each 5 minute period across participant's thigh data.
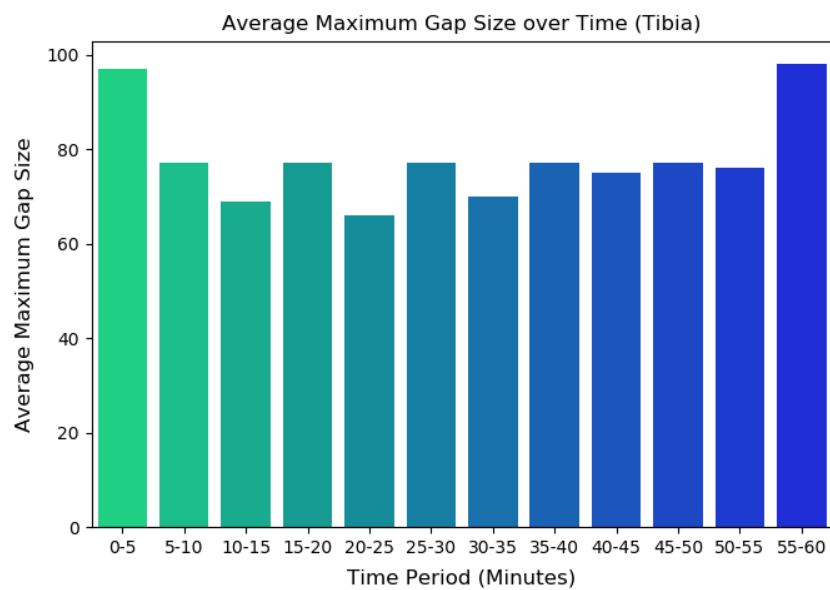
Figure 11.4: The average maximum gap size that occurred across each 5 minute period across participant's tibia data.

Results

The data were analysed to examine the average number of times that data loss occurred across participants in each 5 minute window throughout the recording. Across the tibia, thigh and sacrum,

3
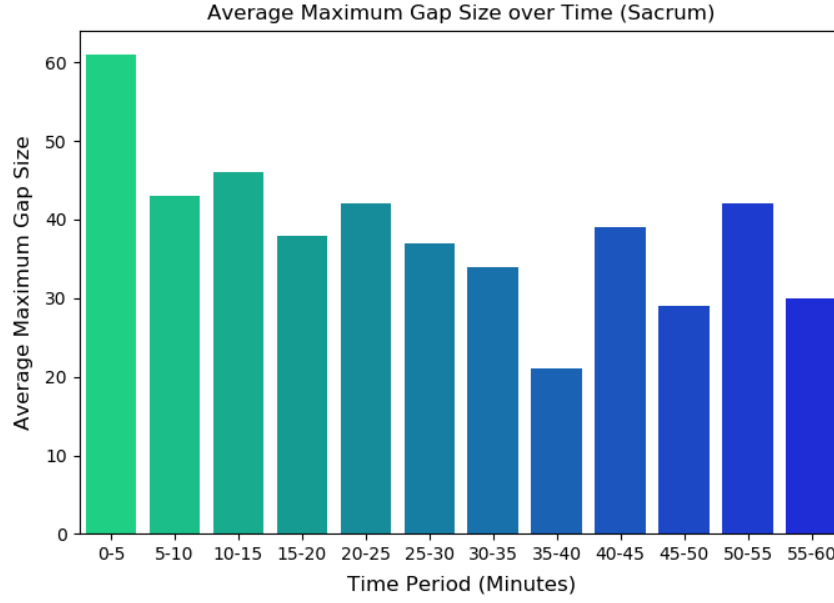
Figure 11.5: The average maximum gap size that occurred across each 5 minute period across participant's sacrum data.
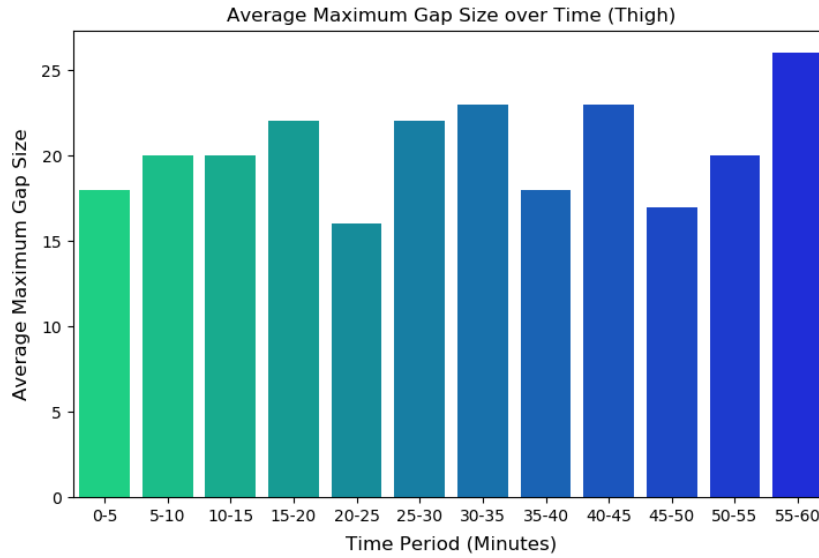
Figure 11.6: The average maximum gap size that occurred across each 5 minute period across participant's thigh data.

no obvious patterns can be seen related to the timing of the data loss. For each sensor/segment, it can be seen in figures 11.1, 11.2 and 11.3 that there was little variation in the number of gaps across the 5 minute periods throughout the recording.

Average maximum gap size loss was calculated for each participant's segment data as the average of the maximum gap sizes across all 5 minute windows for that segment. The average maximum
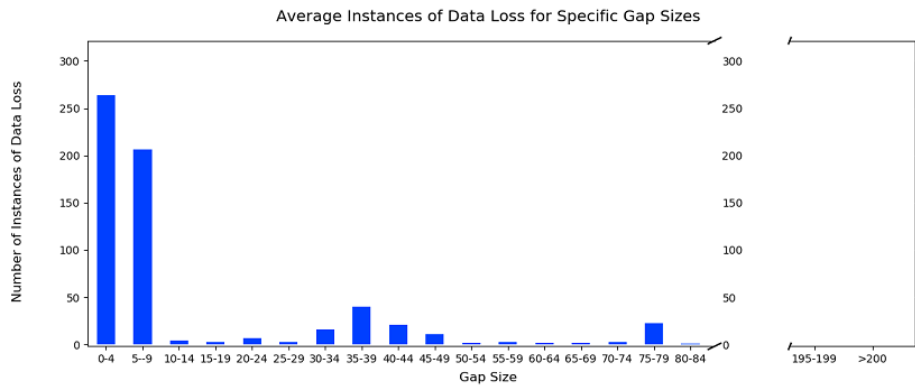
Figure 11.7: The average number of occurrences of gap sizes of specific lengths across participant's tibia data. This was broken into small ranges i.e. 0-4 consecutive missing samples.
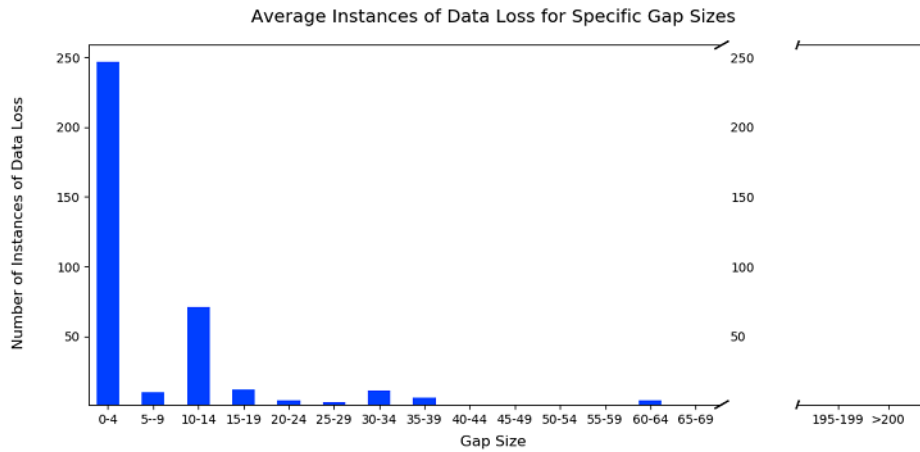
Figure 11.8: The average number of occurrences of gap sizes of specific lengths across participant's sacrum data. This was broken into small ranges i.e. 0-4 consecutive missing samples.
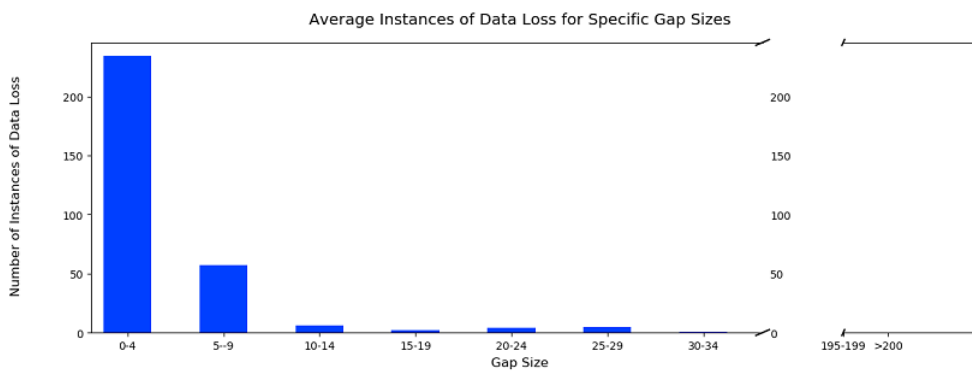
Figure 11.9: The average number of occurrences of gap sizes of specific lengths across participant's thigh data. This was broken into small ranges i.e. 0-4 consecutive missing samples.

gap sizes for the tibia, thigh and sacrum across participants ranged from (66 to 98) samples, (16 to 26) samples and (21 to 61) samples respectively (figures 11.4, 11.5 and 11.6). This tells us that on average, higher amounts of consecutive data was lost in the tibia compared to the thigh and sacrum,

and that the maximum gap size could occur in the range of 80 to 100 samples, and even be as large as > 200 samples in outlier occurrences. These differences could be sensor specific or related to forces and impact magnitudes at these segments.

To indicate whether certain sizes of data loss were more common and if groups of consecutive samples >=200 were being lost, the number of occurrences of fixed gap sizes data losses were examined. The results showed for the tibia, sacrum and thigh data, most of the data loss was in the range of (0 to 4) consecutive samples, closely followed by the range of (5 to 9) consecutive samples. These results can be seen in figures 11.7, 11.9, and 11.8. Very few groups of consecutive sample losses were greater than 80 samples, with 339 samples as the largest number of consecutive samples lost, an outlier which occurred 1 time.

Summary

Using the results which show that the largest average maximum gap sizes were in the range of 80-100 samples from this preliminary test, the thresholds for pre-filling consecutive missing samples were chosen for Section 7.4. The thresholds were set as: 0, <10, <20, <40, <80, <100, <400.

11.2 Appendix B: Data-Loss Additional Experiment

Aim

In order to better explain the finding on why the linear interpolation and previous sample filling had no major improvement on orientation estimation we investigated the factors that were affecting the accuracy of orientation.

Methodology

A 1-minute sample signal of inertial data was extracted from one subject's thigh data where no data loss had occurred and artificial gaps were removed from the data. This was performed separately during fast changing periods of the signal (figure 11.10) and during slow changing periods of the signal (figure 11.11). This was performed twice at each location, once with 10 samples removed from two adjacent strides, and again with 20 samples removed from two adjacent strides. This created 4 inertial signals which were used to calculate orientation and compared to the orientation calculated with the gap free inertial data sample. Before the signals could be compared, the lossy signals were gap filled after orientation was calculated using a linear interpolation to align with the loss-free orientation data.
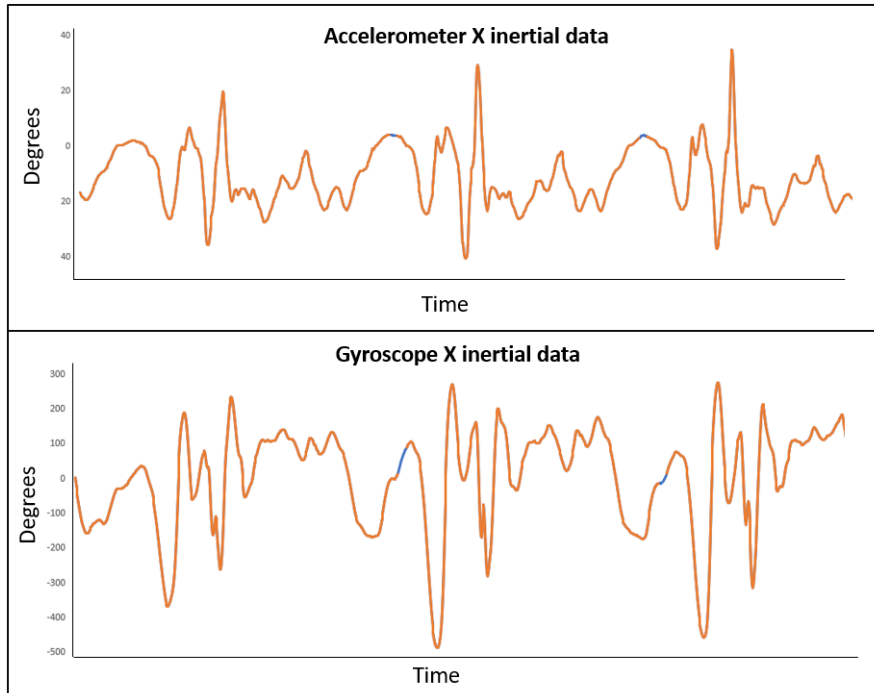
Figure 11.10: The orange signal shows raw thigh segment accelerometer and gyroscope X-axis data, where gaps have been artificially created (blue section) in slow changing regions of the signal.
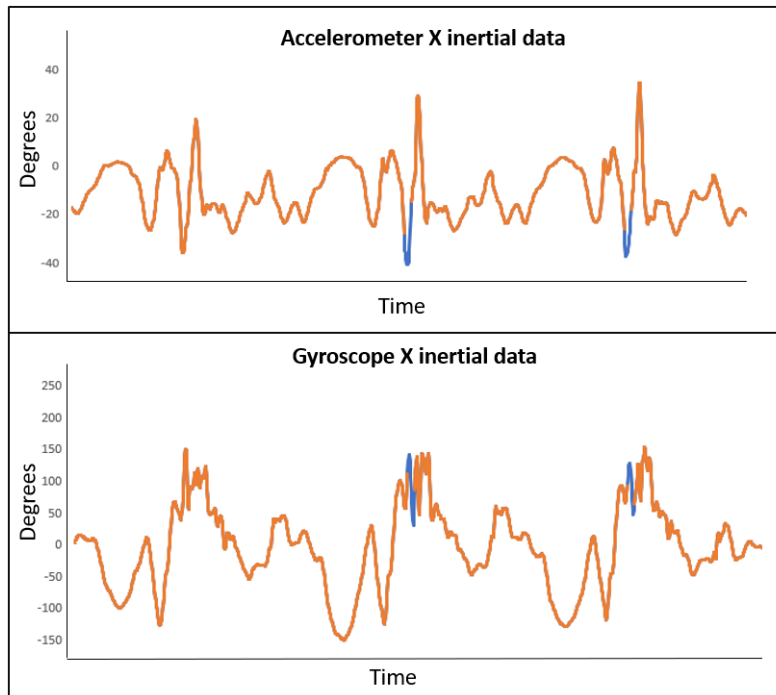
Figure 11.11: The orange signal shows raw thigh segment accelerometer and gyroscope X-axis data, where gaps have been artificially created (blue section) in fast changing regions of the signal.
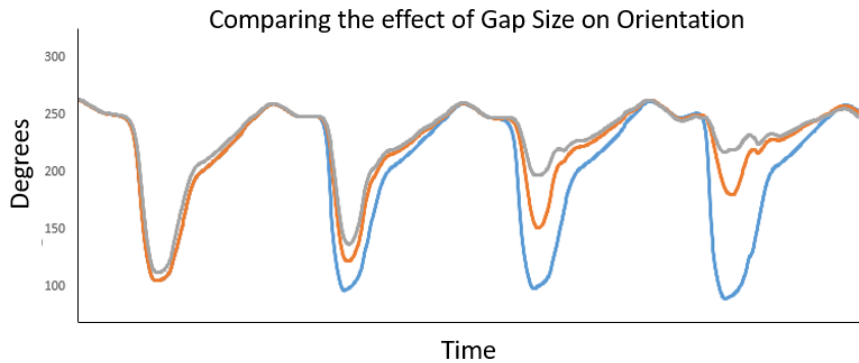
Figure 11.12: The blue signal shows the calculated orientation of the gap-free inertial data. The orange and blue signals show the calculated orientation when 10 samples, and 20 samples, respectively, were removed from slow changing regions of inertial data.

Results

The results of these experiments are shown visually in figures 11.12 and 11.13. The first observation to be made is how the difference in the resulting orientation when data is removed from different regions in the inertial data. This tells us that the resulting orientation is heavily influenced by the region of data where the data loss occurs. Secondly, at each of these two locations, varying the gap size of data removed had a noticeable effect on orientation which suggests that gap size is an influencing factor. It is possible that this effect of gap size was not noticed in our main study 7.4 as the orientation was

2
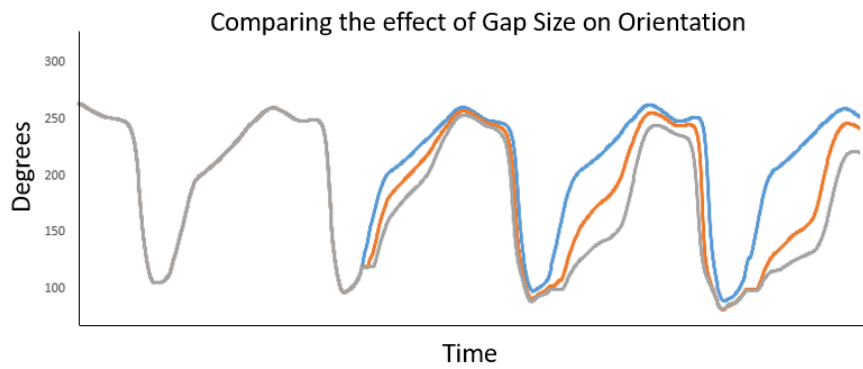
Figure 11.13: The blue signal shows the calculated orientation of the gap-free inertial data. The orange and blue signals show the calculated orientation when 10 samples, and 20 samples, respectively, were removed from fast changing regions of inertial data.

calculated over the entire dataset, and so had time to re-converge after periods of bad data loss before the comparison windows occurred. This would explain why the interpolation performed as well as not filling the gaps in the data.

11.3 Appendix C: Investigating the Benefit of the Magnetometer to Improve Orientation Estimation

Aim

This experiment was initiated to investigate the added benefit of the inclusion of the magnetometer in the Madgwick Filter [14] algorithm implemented for calculating human segment orientation. The experiment was designed to collect inertial sensor data and Vicon optical reference data simultaneously, in order to compare the 9DOF orientation (magnetometer included) calculated by both data capture systems during walking and running. This experiment aimed to quantify to what extent the magnetometer can alleviate some of the problems in the orientation signal experienced in Sections 7.3 and 7.4, and which issues in the signal can be explained by its exclusion.

Methodology

The data in this experiment were captured using one inertial sensor (Shimmer Sensing, Dublin, Ireland) and an optical motion analysis reference system (Vicon Oxford Metrics, Oxford, United Kingdom). The inertial sensor used was a Shimmer3 inertial measurement unit which contains a tri-axial accelerometer, tri-axial gyroscope and tri-axial magnetometer. Inertial data and optical reference data were sampled at 442Hz. The inertial sensor was calibrated using the Shimmer 9DOF Calibration Application (v2.10) and configured for the experiment using Shimmer Consensys Basic (v1.5.4).

Data was collected on the tibia of a single subject in two recordings, first performing a 40 minute walk, followed by a 40 minute run. In each of these recordings an inertial sensor and retro-reflective markers were attached to a plastic box which was taped to the right tibia of the subject. Tri-axial accelerometer, gyroscope and magnetometer data were collected using the inertial sensor and later used to calculate two orientation signals, one from the walk data and one from the run data. The retro-reflective markers attached to the plastic housing were used to calculate orientation using a local coordinate system defined by the markers. The inertial sensor was placed at the origin of this local coordinate system so that the orientation output from the two sets of captured data (Shimmer and Vicon) could be directly compared using RMSE. Walk and run data orientation was calculated using both the Madgwick 6DOF and 9DOF orientation estimation algorithms for comparison.

Results

Firstly, by comparing the 6DOF and 9DOF running orientation visually (figure 11.14), it is clear to see how the errors attributed to fast running in the 6DOF are removed and a smoother estimate is achieved with the 9DOF estimation. This is due to the additional reference data provided by the magnetometer allowing for an absolute estimate of orientation to be achieved. Secondly, and unexpectedly the data loss affects the two signals very differently. In the 6DOF data, inaccuracies which occur before the data loss become amplified after the missing data when the algorithm fails to converge quickly to the correct orientation. Comparing this to the 9DOF data where the signal is relatively smooth before the data loss, and appears only momentarily affected by the missing samples as the algorithm very quickly gets back on track.

(a) X-axis tibia data

(b) Y-axis tibia data
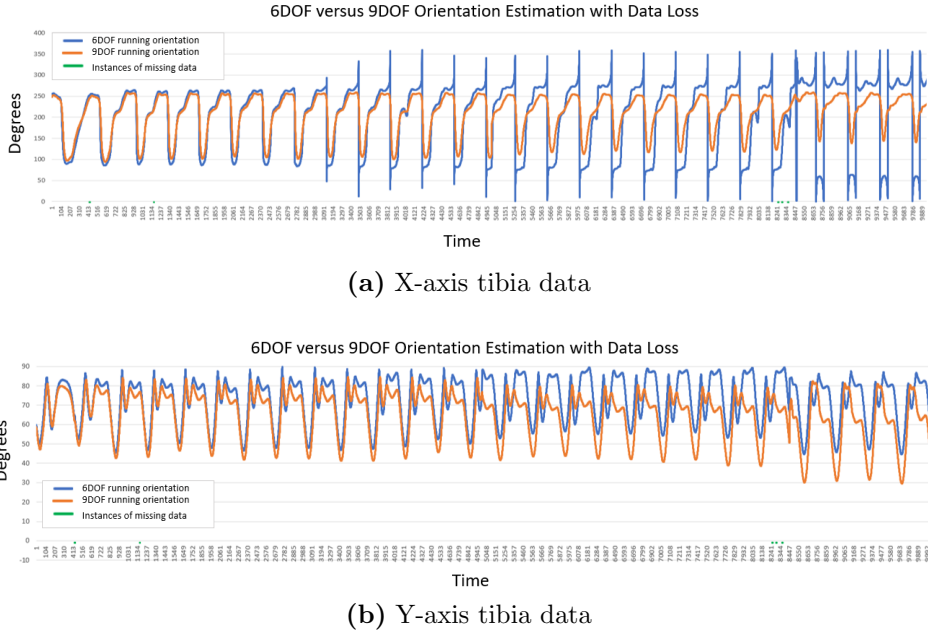
Figure 11.14: Comparison of running orientation estimated using Madgwick 6DOF filter (blue) and Madgwick 9DOF filter (orange) with the presence of data loss (green).

Summary

These short visual experiments show the shortcomings of 6DOF for fast movements compared to 9DOF, which was known in advance of the experiment. In addition, it indicates that the problem being investigated in Section 11.1 is solely a 6DOF orientation estimation problem where data loss compounds the issue of algorithm convergence, and that the additional reference of the magnetometer eliminates this problem entirely.

11.4 Appendix D: Machine Learning Injury prediction - Additional Results

The tables and figures included in this section are in support of the results discussed in chapter 9.

Table 11.1: Approach (a): This table shows the mean score and standard deviations performance of each of the seven classifiers on each of the six datasets. The horizontal blue averages represent the average mean score of each classifier across all datasets, and the vertical blue averages represent the average mean score of each dataset across all classifiers.

Datasets	LogR		NB		KNN		DT		RF		MLP		BaggingSVM		Average:
Approach A	Mean Score	Mean SD	Mean Score	Mean SD	Mean Score	Mean SD	Mean Score	Mean SD	Mean Score	Mean SD	Mean Score	Mean SD	Mean Score	Mean SD	
C	0.56	0.04	0.58	0.08	0.56	0.03	0.55	0.06	0.58	0.10	0.54	0.06	0.53	0.04	0.56
K	0.45	0.07	0.47	0.06	0.41	0.09	0.48	0.14	0.51	0.01	0.47	0.03	0.46	0.08	0.46
I	0.50	0.08	0.54	0.09	0.52	0.07	0.50	0.12	0.43	0.09	0.57	0.05	0.54	0.03	0.51
IC	0.50	0.08	0.54	0.09	0.52	0.07	0.51	0.12	0.46	0.06	0.55	0.06	0.51	0.06	0.51
CK	0.46	0.08	0.48	0.04	0.46	0.08	0.54	0.06	0.50	0.07	0.51	0.06	0.45	0.05	0.49
ICK	0.54	0.08	0.54	0.09	0.53	0.09	0.46	0.05	0.52	0.08	0.53	0.11	0.53	0.09	0.52
Average:	0.50		0.53		0.50		0.51		0.50		0.53		0.50		

Table 11.2: Approach (b): This table shows the mean score and standard deviations performance of each of the seven classifiers on each of the six datasets. The horizontal blue averages represent the average mean score of each classifier across all datasets, and the vertical blue averages represent the average mean score of each dataset across all classifiers.

Datasets	LogR		NB		KNN		DT		RF		MLP		BaggingSVM		Average:
Approach B	Mean Score	Mean SD	Mean Score	Mean SD	Mean Score	Mean SD	Mean Score	Mean SD	Mean Score	Mean SD	Mean Score	Mean SD	Mean Score	Mean SD	
C	0.61	0.06	0.55	0.06	0.51	0.05	0.59	0.05	0.58	0.04	0.58	0.05	0.54	0.03	0.57
K	0.46	0.06	0.49	0.01	0.56	0.03	0.52	0.06	0.50	0.06	0.52	0.07	0.44	0.07	0.50
I	0.45	0.03	0.47	0.03	0.49	0.04	0.50	0.03	0.47	0.05	0.48	0.02	0.44	0.03	0.47
IC	0.48	0.05	0.50	0.03	0.49	0.04	0.47	0.03	0.48	0.06	0.51	0.04	0.48	0.04	0.49
CK	0.53	0.09	0.48	0.04	0.50	0.04	0.54	0.09	0.52	0.08	0.54	0.07	0.51	0.09	0.52
ICK	0.47	0.10	0.48	0.05	0.50	0.05	0.50	0.03	0.48	0.08	0.44	0.06	0.47	0.08	0.48
Average:	0.50		0.50		0.51		0.52		0.51		0.51		0.48		

DATASET- C	Precision	Recall	F1	CM (tn, fp, fn, tp):
LogR	0.56	0.51	0.53	9.4 6.0 7.0 7.4
NB	0.45	0.63	0.52	8.2 7.2 5.4 9.0
KNN	0.55	0.53	0.54	9.2 6.2 6.8 7.6
DT	0.54	0.51	0.52	9.0 6.4 7.0 7.4
RF	0.60	0.50	0.53	10.0 5.4 7.2 7.2
MLP	0.49	0.29	0.34	12.0 3.4 10.2 4.2
BaggingSVM	0.64	0.35	0.37	10.8 4.6 9.4 5.0

DATASET- IC	Precision	Recall	F1	CM (tn, fp, fn, tp):
LogR	0.50	0.33	0.39	10.2 5.2 9.6 4.8
NB	0.46	0.30	0.36	12.0 3.4 10.0 4.4
KNN	0.50	0.40	0.43	10.0 5.4 8.6 5.8
DT	0.51	0.47	0.47	8.4 7.0 7.6 6.8
RF	0.45	0.40	0.42	8.0 7.4 8.6 5.8
MLP	0.64	0.42	0.45	10.4 5.0 8.4 6.0
BaggingSVM	0.49	0.60	0.50	6.6 8.8 5.8 8.6

DATASET- K	Precision	Recall	F1	CM (tn, fp, fn, tp):
LogR	0.39	0.33	0.35	8.6 6.8 9.6 4.8
NB	0.45	0.61	0.51	5.2 10.2 5.6 8.8
KNN	0.38	0.35	0.36	7.2 8.2 9.4 5.0
DT	0.48	0.39	0.43	8.8 6.6 8.8 5.6
RF	0.49	0.45	0.46	8.8 6.6 8.0 6.4
MLP	0.34	0.51	0.38	6.4 9.0 7.2 7.2
BaggingSVM	0.44	0.60	0.50	5.0 10.4 5.8 8.6

DATASET- CK	Precision	Recall	F1	CM (tn, fp, fn, tp):
LogR	0.42	0.35	0.38	8.8 6.6 9.4 5.0
NB	0.36	0.46	0.41	7.6 7.8 7.8 6.6
KNN	0.44	0.42	0.43	7.8 7.6 8.4 6.0
DT	0.55	0.53	0.51	8.6 6.8 6.8 7.6
RF	0.48	0.51	0.49	7.6 7.8 7.0 7.4
MLP	0.50	0.63	0.54	6.0 9.4 5.4 9.0
BaggingSVM	0.42	0.42	0.41	7.4 8.0 8.4 6.0

DATASET- I	Precision	Recall	F1	CM (tn, fp, fn, tp):
LogR	0.50	0.33	0.39	10.2 5.2 9.6 4.8
NB	0.46	0.30	0.36	12.0 3.4 10.0 4.4
KNN	0.50	0.39	0.42	10.2 5.2 8.8 5.6
DT	0.51	0.44	0.47	8.6 6.8 8.0 6.4
RF	0.43	0.39	0.43	7.4 8.0 8.8 5.6
MLP	0.56	0.33	0.37	12.4 3.0 9.6 4.8
BaggingSVM	0.54	0.64	0.53	6.8 8.6 5.2 9.2

DATASET- ICK	Precision	Recall	F1	CM (tn, fp, fn, tp):
LogR	0.55	0.36	0..43	11.0 4.4 9.2 5.2
NB	0.46	0.30	0.36	12.0 3.4 10.0 4.4
KNN	0.53	0.39	0.43	10.4 5.0 8.8 5.6
DT	0.45	0.40	0.40	8.0 7.4 8.6 5.8
RF	0.52	0.47	0.48	8.6 6.8 7.6 6.8
MLP	0.60	0.60	0.54	7.2 8.2 5.8 8.6
BaggingSVM	0.48	0.75	0.56	4.6 10.8 3.6 10.8

Figure 11.15: Approach (a): Details of the precision, recall, F1 scores and Confusion Matrices (CM) for each combination of classifier and dataset, where CM (tn, fp, fn, tp) refers to the confusion matrix with values (True Negative, False Positive, False Negative, True Positive).

3

DATASET - C	Precision	Recall	F1	CM (tn, fp, fn, tp):
LogR	0.60	0.61	0.61	762.4 494.4 478.4 741.6
NB	0.43	0.60	0.50	637.4 619.4 491.4 728.6
KNN	0.40	0.40	0.40	766.4 490.4 734.0 486.0
DT	0.60	0.40	0.49	946.0 310.8 707.6 512.4
RF	0.62	0.42	0.50	924.4 332.4 704.0 516.0
MLP	0.57	0.67	0.60	629.8 627.0 408.8 811.2
BaggingSVM	0.53	0.47	0.50	749.8 507.0 645.0 575.0

DATASET - IC	Precision	Recall	F1	CM (tn, fp, fn, tp):
LogR	0.47	0.38	0.42	727.0 529.8 755.2 464.8
NB	0.2	0.2	0.2	1019.6 237.2 971.6 248.4
KNN	0.47	0.54	0.49	552.6 704.2 559.4 660.6
DT	0.46	0.39	0.41	701.4 555.4 739.4 480.6
RF	0.47	0.42	0.44	686.4 570.4 714.2 505.8
MLP	0.41	0.35	0.38	829.8 427.0 791.2 428.8
BaggingSVM	0.47	0.35	0.4	776.0 480.8 797.0 423.0

DATASET - K	Precision	Recall	F1	CM (tn, fp, fn, tp):
LogR	0.45	0.40	0.43	645.6 611.2 711.2 508.8
NB	0.00	0.00	0.00	1241.4 15.4 1220.0 0.0
KNN	0.55	0.58	0.57	686.8 570.0 505.4 714.6
DT	0.52	0.39	0.43	826.4 430.4 740.2 479.8
RF	0.50	0.44	0.46	714.8 542.0 683.0 537.0
MLP	0.51	0.53	0.52	639.4 617.4 570.6
BaggingSVM	0.44	0.41	0.42	624.0 632.8 725.6 494.4

DATASET - CK	Precision	Recall	F1	CM (tn, fp, fn, tp):
LogR	0.52	0.53	0.52	677.4 579.4 574.8 645.2
NB	0.08	0.09	0.09	1093.4 163.4 1109.8 110.2
KNN	0.49	0.46	0.47	678.6 578.2 664.2 555.8
DT	0.53	0.55	0.53	676.8 580.0 552.4 667.6
RF	0.51	0.48	0.49	702.8 554.0 627.0 593.0
MLP	0.54	0.57	0.55	644.6 612.2 520.0 700.0
BaggingSVM	0.51	0.44	0.46	741.2 515.6 690.8 529.2

DATASET - I	Precision	Recall	F1	CM (tn, fp, fn, tp):
LogR	0.42	0.30	0.34	754.4 502.4 861.8 358.2
NB	0.26	0.23	0.24	898.4 358.4 943.4 276.6
KNN	0.48	0.42	0.44	716.4 540.4 712.0 508.0
DT	0.49	0.48	0.48	644.0 612.8 628.8 591.2
RF	0.47	0.50	0.47	575.4 681.4 610.2 609.8
MLP	0.18	0.16	0.16	994.0 262.8 1035.4 184.6
BaggingSVM	0.42	0.33	0.37	707.2 549.6 816.6 403.4

DATASET - ICK	Precision	Recall	F1	CM (tn, fp, fn, tp):
LogR	0.46	0.44	0.46	643.6 613.2 685.2 534.8
NB	0.23	0.17	0.20	970.8 286.0 1002.8 217.2
KNN	0.50	0.60	0.53	511.8 745.0 477.8 742.2
DT	0.48	0.48	0.48	646.4 610.4 631.6 588.4
RF	0.46	0.41	0.43	689.8 567.0 718.8 501.2
MLP	0.32	0.33	0.31	677.2 579.6 801.4 418.6
BaggingSVM	0.46	0.42	0.44	655.6 601.2 703.8 516.2

Figure 11.16: Approach (b): Details of the precision, recall, F1 scores and Confusion Matrices (CM) for each combination of classifier and dataset, where CM (tn, fp, fn, tp) refers to the confusion matrix with values (True Negative, False Positive, False Negative, True Positive).

11.5 Appendix E: Machine Learning Injury prediction using Default Hyper-parameters and Top 20 Fixed Features.

The work in this section was carried out as part of Section 9 and excluded from the main chapter for brevity. This section only includes the results of the machine learning classifiers trained using the default hyper-parameters.

Methodology

The set-up of the models and feature extraction are identical to Study 9, where approach (a) and (b) describe a sample per subject and a sample per stride approach, respectively. For each approach 'a' and 'b', classifiers were trained with different subsets of the dataset: inertial features only (I), kinematics features only (K), clinical data only (C), and combinations of IC, CK and ICK.

Everything remains the same in this methodology except only default hyperparameters are used in the classifiers with the top 20 selected features (as chose by *SelectKbest*). There is no nested cross validation, only one single 5-fold cross validation which performs the same train/test splits as in section 9.

The results for the default parameters classifiers (K-Nearest Neighbour, Decision Tree, Logistic Regression, Random Forest, Naive Bayes, Bagged Support Vector Machine and Multi-Layer Perceptron) can be seen in the figures below.

Results

The AUC scores across classifiers and datasets.

Table 11.3: Approach (a): This table shows the mean score and standard deviations performance of each of the seven classifiers on each of the six datasets. The horizontal blue averages represent the average mean score of each classifier across all datasets, and the vertical blue averages represent the average mean score of each dataset across all classifiers.

Datasets	LogR		NB		KNN		DT		RF		MLP		BaggingSVM		Average:
Approach A	Mean Score	Mean SD	Mean Score	Mean SD	Mean Score	Mean SD	Mean Score	Mean SD	Mean Score	Mean SD	Mean Score	Mean SD	Mean Score	Mean SD	
C	0.56	0.04	0.55	0.09	0.62	0.11	0.46	0.06	0.52	0.04	0.55	0.06	0.55	0.09	0.54
K	0.46	0.08	0.44	0.06	0.45	0.06	0.44	0.09	0.44	0.05	0.43	0.07	0.41	0.08	0.44
I	0.52	0.08	0.43	0.07	0.48	0.04	0.50	0.09	0.50	0.10	0.49	0.05	0.50	0.09	0.49
IC	0.51	0.07	0.42	0.06	0.49	0.05	0.48	0.06	0.50	0.10	0.49	0.05	0.51	0.09	0.49
CK	0.43	0.05	0.43	0.06	0.45	0.06	0.44	0.07	0.43	0.06	0.43	0.06	0.40	0.07	0.43
ICK	0.48	0.07	0.39	0.05	0.43	0.08	0.51	0.10	0.47	0.09	0.48	0.10	0.46	0.08	0.46
Average:	0.48		0.42		0.46		0.47		0.47		0.46		0.46		

Table 11.4: Approach (b): This figure shows the mean score and standard deviations performance of each of the seven classifiers on each of the six datasets. The horizontal blue averages represent the average mean score of each classifier across all datasets, and the vertical blue averages represent the average mean score of each dataset across all classifiers.

Datasets	LogR		NB		KNN		DT		RF		MLP		BaggingSVM		Average:
Approach B	Mean Score	Mean SD	Mean Score	Mean SD	Mean Score	Mean SD	Mean Score	Mean SD	Mean Score	Mean SD	Mean Score	Mean SD	Mean Score	Mean SD	
C	0.58	0.04	0.61	0.04	0.48	0.07	0.51	0.09	0.57	0.07	0.53	0.04	0.55	0.05	0.55
K	0.49	0.06	0.47	0.06	0.57	0.04	0.47	0.05	0.53	0.06	0.51	0.05	0.48	0.07	0.50
I	0.49	0.10	0.51	0.12	0.52	0.05	0.50	0.03	0.52	0.08	0.50	0.05	0.49	0.10	0.50
IC	0.50	0.09	0.47	0.11	0.50	0.09	0.49	0.05	0.49	0.05	0.51	0.08	0.51	0.10	0.50
CK	0.52	0.06	0.55	0.08	0.53	0.05	0.55	0.08	0.56	0.11	0.50	0.08	0.52	0.06	0.53
ICK	0.48	0.09	0.48	0.11	0.53	0.04	0.51	0.07	0.50	0.07	0.51	0.07	0.47	0.11	0.50
Average:	0.50		0.50		0.53		0.50		0.52		0.51		0.49		

DATASET - C	Precision	Recall	F1	CM (tn, fp, fn, tp):
LogR	0.54	0.55	0.54	8.8 6.6 6.4 8.0
NB	0.52	0.70	0.59	6.2 9.2 4.4 10.0
KNN	0.60	0.60	0.60	9.8 5.6 5.8 8.6
DT	0.44	0.38	0.40	8.4 7.0 9.0 5.4
RF	0.51	0.53	0.51	8.0 7.4 6.8 7.6
MLP	0.53	0.51	0.52	9.0 6.4 7.0 7.4
BaggingSVM	0.54	0.39	0.44	11.0 4.4 8.8 5.6

DATASET - IC	Precision	Recall	F1	CM (tn, fp, fn, tp):
LogR	0.48	0.46	0.47	8.6 6.8 7.8 6.6
NB	0.38	0.39	0.37	7.2 8.2 8.8 5.6
KNN	0.46	0.39	0.42	9.0 6.4 8.8 5.6
DT	0.45	0.46	0.45	7.6 7.8 7.8 6.6
RF	0.46	0.43	0.44	8.6 6.8 8.2 6.2
MLP	0.47	0.43	0.45	8.6 6.8 8.2 6.2
BaggingSVM	0.49	0.43	0.46	9.0 6.4 8.2 6.2

DATASET - K	Precision	Recall	F1	CM (tn, fp, fn, tp):
LogR	0.44	0.39	0.41	8.2 7.2 8.8 5.6
NB	0.43	0.45	0.43	6.6 8.8 8.0 6.4
KNN	0.43	0.46	0.44	6.8 8.6 7.8 6.6
DT	0.43	0.42	0.42	7.2 8.2 8.4 6.0
RF	0.42	0.39	0.4	7.6 7.8 8.8 5.6
MLP	0.37	0.33	0.34	8.0 7.4 9.6 4.8
BaggingSVM	0.28	0.36	0.31	8.2 7.2 10.4 4.0

DATASET - CK	Precision	Recall	F1	CM (tn, fp, fn, tp):
LogR	0.4	0.34	0.36	8.0 7.4 9.6 4.8
NB	0.41	0.45	0.42	6.4 9.0 8.0 6.4
KNN	0.44	0.47	0.45	6.6 8.8 7.6 6.8
DT	0.42	0.43	0.42	6.8 8.6 8.2 6.2
RF	0.41	0.36	0.38	7.8 7.6 9.2 5.2
MLP	0.39	0.35	0.37	7.8 7.6 9.4 5.0
BaggingSVM	0.36	0.31	0.33	7.6 7.8 10.0 4.4

DATASET - I	Precision	Recall	F1	CM (tn, fp, fn, tp):
LogR	0.49	0.46	0.47	8.8 6.6 7.8 6.6
NB	0.39	0.36	0.36	7.8 7.6 9.2 5.2
KNN	0.46	0.38	0.41	9.0 6.4 9.0 5.4
DT	0.47	0.46	0.46	8.2 7.2 7.8 6.6
RF	0.46	0.43	0.44	8.6 6.8 8.2 6.2
MLP	0.47	0.43	0.45	8.6 6.8 8.2 6.2
BaggingSVM	0.48	0.42	0.45	9.0 6.4 8.4 6.0

DATASET - ICK	Precision	Recall	F1	CM (tn, fp, fn, tp):
LogR	0.47	0.45	0.45	8.0 7.4 8.0 6.4
NB	0.34	0.38	0.33	6.8 8.6 9.4 5.0
KNN	0.4	0.38	0.38	7.6 7.8 9.0 5.4
DT	0.48	0.46	0.47	8.6 6.8 7.8 6.6
RF	0.42	0.37	0.4	8.6 6.8 9.0 5.4
MLP	0.46	0.43	0.44	8.2 7.2 8.2 6.2
BaggingSVM	0.44	0.43	0.43	7.6 7.8 8.2 6.2

Figure 11.17: Approach (a): Details of the precision, recall, F1 scores and Confusion Matrices (CM) for each combination of classifier and dataset, where CM (tn, fp, fn, tp) refers to the confusion matrix with values (True Negative, False Positive, False Negative, True Positive).

3

DATASET - C	Precision	Recall	F1	CM (tn, fp, fn, tp):
LogR	0.57	0.58	0.57	: 709.0 547.8 508.2 711.8
NB	0.57	0.79	0.66	521.6 735.2 251.8 968.2
KNN	0.46	0.44	0.44	: 649.8 607.0 683.2 536.8
DT	0.50	0.52	0.51	628.0 628.8 584.2 635.8
RF	0.54	0.57	0.55	753.8 503.0 564.0 656.0
MLP	0.53	0.50	0.50	695.6 561.2 615.8 604.2
BaggingSVM	0.54	0.49	0.51	749.8 507.0 628.2 591.8

DATASET - IC	Precision	Recall	F1	CM (tn, fp, fn, tp):
LogR	0.49	0.48	0.48	663.0 593.8 635.8 584.2
NB	0.42	0.43	0.42	643.4 613.4 689.8 530.2
KNN	0.48	0.46	0.47	680.6 576.2 657.8 562.2
DT	0.48	0.48	0.48	620.0 636.8 633.6 586.4
RF	0.47	0.43	0.45	678.2 578.6 695.6 524.4
MLP	0.50	0.52	0.51	630.6 626.2 586.0 634.0
BaggingSVM	0.50	0.47	0.49	694.6 562.2 645.0 575.0

DATASET - K	Precision	Recall	F1	CM (tn, fp, fn, tp):
LogR	0.49	0.47	0.48	633.2 623.6 640.8 579.2
NB	0.46	0.45	0.45	625.4 631.4 669.0 551.0
KNN	0.56	0.59	0.57	691.6 565.2 496.4 723.6
DT	0.46	0.41	0.43	659.6 597.2 712.6 507.4
RF	0.54	0.52	0.52	693.2 563.6 585.6 634.4
MLP	0.50	0.53	0.51	605.4 651.4 570.2 649.8
BaggingSVM	0.48	0.43	0.45	652.0 604.8 690.8 529.2

DATASET - CK	Precision	Recall	F1	CM (tn, fp, fn, tp):
LogR	0.52	0.53	0.52	654.2 602.6 572.0 648.0
NB	0.54	0.61	0.57	621.0 635.8 471.8 748.2
KNN	0.52	0.55	0.53	635.8 621.0 541.2 678.8
DT	0.55	0.57	0.56	675.0 581.8 529.2 690.8
RF	0.57	0.55	0.55	731.2 525.6 551.0 669.0
MLP	0.50	0.49	0.49	654.0 602.8 623.8 596.2
BaggingSVM	0.51	0.53	0.52	630.8 626.0 566.2 653.8

DATASET - I	Precision	Recall	F1	CM (tn, fp, fn, tp):
LogR	0.48	0.44	0.46	672.4 584.4 676.8 543.2
NB	0.45	0.39	0.41	802.4 454.4 742.6 477.4
KNN	0.51	0.45	0.47	722.0 534.8 660.6 559.4
DT	0.50	0.45	0.47	703.4 553.4 673.0 547.0
RF	0.51	0.42	0.47	772.6 484.2 697.0 523.0
MLP	0.49	0.48	0.48	664.8 592.0 639.6 580.4
BaggingSVM	0.47	0.43	0.45	685.4 571.4 694.8 525.2

DATASET - CK	Precision	Recall	F1	CM (tn, fp, fn, tp):
LogR	0.47	0.44	0.45	658.8 598.0 686.8 533.2
NB	0.41	0.43	0.42	673.0 583.8 681.6 538.4
KNN	0.53	0.53	0.52	680.0 576.8 576.8 643.2
DT	0.50	0.56	0.53	563.4 693.4 530.0 690.0
RF	0.50	0.49	0.49	658.2 598.6 620.2 599.8
MLP	0.51	0.50	0.51	645.4 611.4 602.6 617.4
BaggingSVM	0.47	0.43	0.45	660.4 596.4 699.2 520.8

Figure 11.18: Approach (b): Details of the precision, recall, F1 scores and Confusion Matrices (CM) for each combination of classifier and dataset, where CM (tn, fp, fn, tp) refers to the confusion matrix with values (True Negative, False Positive, False Negative, True Positive).

11.6   Appendix F: Machine Learning Injury prediction using Tuned Hyper-parameters with Top 20 Fixed Features.

The work in this section was carried out as part of Section 9 and excluded from the main chapter for brevity. This section only includes the results of the machine learning classifiers trained using the tuned hyper-parameters with top 20 fixed features selected.

Methodology

The set-up of the models and feature extraction are identical to Study 9, where approach (a) and (b) describe a sample per subject and a sample per stride approach, respectively. For each approach 'a' and 'b', classifiers were trained with different subsets of the dataset: inertial features only (I), kinematics features only (K), clinical data only (C), and combinations of IC, CK and ICK.

Everything remains the same in this methodology except the classifiers have tuned hyper-parameters and use a fixed set of top 20 fixed features (as selected by SelectKbest from the data). nested cross validation is used where the inner loop is responsible for tuning the hyperparameters and the outer loop evaluates the model with the test data using the same train/test splits as in section 9.

The results for the tuned hyper-parameters with top 20 fixed feature classifiers (K-Nearest Neighbour, Decision Tree, Logistic Regression, Random Forest, Naive Bayes, Bagged Support Vector Machine and Multi-Layer Perceptron) can be seen in the figures below.

Results

The AUC scores across classifiers and datasets.

Datasets	LogR		NB		KNN		DT		RF		MLP		BaggingSVM		
Approach A	Mean Score	Mean SD	Mean Score	Mean SD	Mean Score	Mean SD	Mean Score	Mean SD	Mean Score	Mean SD	Mean Score	Mean SD	Mean Score	Mean SD	Average:
C	0.55	0.05	0.56	0.08	0.51	0.04	0.45	0.06	0.54	0.08	0.50	0.00	0.47	0.08	0.51
K	0.45	0.06	0.48	0.07	0.44	0.06	0.48	0.12	0.48	0.04	0.47	0.06	0.49	0.06	0.47
I	0.47	0.05	0.51	0.02	0.49	0.10	0.45	0.04	0.45	0.06	0.48	0.04	0.53	0.05	0.48
IC	0.46	0.06	0.44	0.06	0.50	0.09	0.47	0.05	0.53	0.05	0.48	0.07	0.49	0.03	0.48
CK	0.44	0.05	0.48	0.08	0.43	0.09	0.49	0.08	0.51	0.08	0.47	0.05	0.46	0.06	0.47
ICK	0.44	0.50	0.43	0.06	0.42	0.05	0.46	0.06	0.53	0.06	0.51	0.04	0.52	0.05	0.47
Average:	0.47		0.49		0.46		0.48		0.50		0.49		0.49		

Figure 11.19: Approach (a): This figure shows the mean score and standard deviations performance of each of the seven classifiers on each of the six datasets. The horizontal blue averages represent the average mean score of each classifier across all datasets, and the vertical blue averages represent the average mean score of each dataset across all classifiers.

Datasets	LogR		NB		KNN		DT		RF		MLP		BaggingSVM		
Approach B	Mean Score	Mean SD	Mean Score	Mean SD	Mean Score	Mean SD	Mean Score	Mean SD	Mean Score	Mean SD	Mean Score	Mean SD	Mean Score	Mean SD	Average:
C	0.58	0.02	0.55	0.07	0.49	0.03	0.58	0.07	0.54	0.08	0.59	0.07	0.56	0.04	0.56
K	0.49	0.07	0.50	0.05	0.58	0.04	0.50	0.04	0.53	0.03	0.51	0.08	0.47	0.07	0.51
I	0.49	0.10	0.50	0.00	0.49	0.07	0.49	0.07	0.48	0.08	0.48	0.04	0.49	0.10	0.49
IC	0.50	0.09	0.52	0.07	0.51	0.09	0.49	0.03	0.51	0.04	0.50	0.08	0.50	0.10	0.50
CK	0.52	0.07	0.48	0.03	0.53	0.05	0.53	0.08	0.49	0.02	0.55	0.06	0.54	0.06	0.52
ICK	0.48	0.10	0.52	0.05	0.54	0.02	0.50	0.07	0.50	0.09	0.49	0.08	0.49	0.09	0.50
Average:	0.51		0.51		0.52		0.52		0.51		0.52		0.51		

Figure 11.20: Approach (b): This figure shows the mean score and standard deviations performance of each of the seven classifiers on each of the six datasets. The horizontal blue averages represent the average mean score of each classifier across all datasets, and the vertical blue averages represent the average mean score of each dataset across all classifiers.

2

DATASET-C	Precision	Recall	F1	CM (tn, fp, fn, tp):
LogR	0.55	0.49	0.51	9.6 5.8 7.4 7.0
NB	0.32	0.46	0.38	10.0 5.4 8.0 6.4
KNN	0.49	0.50	0.49	7.8 7.6 7.2 7.2
DT	0.44	0.41	0.42	7.8 7.6 8.6 5.8
RF	0.52	0.49	0.49	9.0 6.4 7.4 7.0
MLP	0.19	0.19	0.18	12.4 3.0 11.8 2.6
BaggingSVM	0.47	0.44	0.40	8.0 7.4 8.0 6.4

DATASET-IC	Precision	Recall	F1	CM (tn, fp, fn, tp):
LogR	0.44	0.40	0.42	8.0 7.4 8.6 5.8
NB	0.23	0.22	0.23	10.2 5.2 11.2 3.2
KNN	0.52	0.41	0.45	9.4 6.0 8.6 5.8
DT	0.44	0.38	0.40	8.8 6.6 9.0 5.4
RF	0.52	0.46	0.49	9.4 6.0 7.8 6.6
MLP	0.45	0.72	0.54	3.8 11.6 4.0 10.4
BaggingSVM	0.47	0.54	0.48	6.8 8.6 6.7 7.8

DATASET-K	Precision	Recall	F1	CM (tn, fp, fn, tp):
LogR	0.40	0.33	0.35	8.6 6.8 9.6 4.8
NB	0.46	0.29	0.33	10.2 5.2 10.2 4.2
KNN	0.41	0.32	0.35	8.6 6.8 9.8 4.6
DT	0.46	0.44	0.45	8.0 7.4 8.0 6.4
RF	0.46	0.49	0.47	7.2 8.2 7.4 7.0
MLP	0.41	0.37	0.38	8.6 6.8 9.0 5.4
BaggingSVM	0.46	0.68	0.52	4.6 10.8 4.6 9.8

DATASET-CK	Precision	Recall	F1	CM (tn, fp, fn, tp):
LogR	0.39	0.33	0.35	8.4 7.0 9.6 4.8
NB	0.56	0.33	0.34	9.6 5.8 9.6 4.8
KNN	0.42	0.40	0.41	7.0 8.4 8.6 5.8
DT	0.47	0.41	0.44	8.8 6.6 8.4 6.0
RF	0.50	0.51	0.50	7.6 7.8 7.0 7.4
MLP	0.44	0.63	0.50	4.6 10.8 5.2 9.2
BaggingSVM	0.43	0.52	0.47	6.2 9.2 6.8 7.6

DATASET-I	Precision	Recall	F1	CM (tn, fp, fn, tp):
LogR	0.44	0.42	0.43	8.0 7.4 8.4 6.0
NB	0.21	0.16	0.16	13.4 2.0 12.2 2.2
KNN	0.51	0.41	0.44	9.0 6.4 8.6 5.8
DT	0.40	0.29	0.33	9.2 6.2 10.2 4.2
RF	0.44	0.39	0.41	8.0 7.4 8.8 5.6
MLP	0.46	0.64	0.51	5.2 10.2 5.2 9.2
BaggingSVM	0.51	0.61	0.54	7.0 8.4 5.6 8.8

DATASET-ICK	Precision	Recall	F1	CM (tn, fp, fn, tp):
LogR	0.42	0.40	0.42	7.4 8.0 8.6 5.8
NB	0.34	0.33	0.30	8.8 6.6 10.0 4.4
KNN	0.37	0.25	0.29	9.0 6.4 10.8 3.6
DT	0.44	0.49	0.45	6.8 8.6 7.4 7.0
RF	0.53	0.40	0.44	10.2 5.2 8.6 5.8
MLP	0.49	0.88	0.61	2.2 13.2 1.8 12.6
BaggingSVM	0.51	0.53	0.51	8.0 7.4 6.8 7.6

Figure 11.21: Approach (a): Details of the precision, recall, F1 scores and Confusion Matrices (CM) for each combination of classifier and dataset, where CM (tn, fp, fn, tp) refers to the confusion matrix with values (True Negative, False Positive, False Negative, True Positive).

DATASET-C	Precision	Recall	F1	CM (tn, fp, fn, tp):
LogR	0.57	0.58	0.57	712.2 544.6 508.2 711.8
NB	0.44	0.54	0.48	704.8 552.0 560.2
KNN	0.48	0.46	0.47	646.8 610.0 654.2 565.8
DT	0.56	0.44	0.47	887.2 369.6 685.4
RF	0.51	0.40	0.44	841.4 415.4 730.0 490.0
MLP	0.52	0.41	0.45	951.8 305.0 723.0 497.0
BaggingSVM	0.56	0.50	0.53	779.8 477.0 607.8 612.2

DATASET-IC	Precision	Recall	F1	CM (tn, fp, fn, tp):
LogR	0.50	0.48	0.49	665.4 591.4 635.2 584.8
NB	0.23	0.20	0.21	1071.8 185.0 972.6 247.4
KNN	0.49	0.47	0.47	686.6 570.2 651.2 568.8
DT	0.48	0.46	0.47	661.4 595.4 662.6 557.4
RF	0.51	0.43	0.47	746.2 510.6 690.8 529.2
MLP	0.39	0.39	0.39	754.4 502.4 740.6 479.4
BaggingSVM	0.50	0.45	0.48	708.0 548.8 666.4 553.6

DATASET-K	Precision	Recall	F1	CM (tn, fp, fn, tp):
LogR	0.49	0.48	0.48	643.6 613.2 635.6 584.4
NB	0.19	0.16	0.17	1063.6 193.2 1027.0 193.0
KNN	0.56	0.60	0.58	694.0 562.8 479.0 741.0
DT	0.50	0.44	0.46	708.0 548.8 679.8 540.2
RF	0.53	0.41	0.45	818.0 438.8 727.0 493.0
MLP	0.41	0.48	0.43	685.0 571.8 646.2 573.8
BaggingSVM	0.48	0.42	0.44	672.6 584.2 710.8 509.2

DATASET-CK	Precision	Recall	F1	CM (tn, fp, fn, tp):
LogR	0.51	0.53	0.20	643.8 613.0 572.4 647.6
NB	0.08	0.09	0.08	1100.6 156.2 1107.4 112.6
KNN	0.52	0.55	0.53	638.6 618.2 543.2 676.8
DT	0.50	0.36	0.40	869.4 387.4 786.2 433.8
RF	0.48	0.39	0.42	747.8 509.0 751.4 468.6
MLP	0.54	0.54	0.52	702.8 554.0 559.4 660.6
BaggingSVM	0.55	0.57	0.55	649.6 607.2 524.4 695.6

DATASET-I	Precision	Recall	F1	CM (tn, fp, fn, tp):
LogR	0.48	0.43	0.45	699.4 557.4 694.0 526.0
NB	0.46	0.28	0.29	852.2 404.6 885.8 334.2
KNN	0.47	0.42	0.44	664.0 592.8 701.8 518.2
DT	0.50	0.37	0.42	759.8 497.0 765.2 454.8
RF	0.58	0.40	0.41	732.0 524.8 734.6 485.4
MLP	0.00	0.00	0.00	1256.8 0.0 1220.0 0.0
BaggingSVM	0.48	0.45	0.46	674.4 582.4 676.6 543.4

DATASET-ICK	Precision	Recall	F1	CM (tn, fp, fn, tp):
LogR	0.48	0.44	0.46	650.6 606.2 675.6 544.4
NB	0.12	0.15	0.13	1132.4 124.4 1031.4 188.6
KNN	0.53	0.51	0.52	709.2 547.6 596.8 623.2
DT	0.48	0.41	0.42	723.6 533.2 719.2 500.8
RF	0.47	0.43	0.45	707.6 549.2 692.0 528.0
MLP	0.49	0.48	0.48	632.2 624.6 638.4 581.6
BaggingSVM	0.48	0.44	0.46	666.4 590.4 680.6 539.4

Figure 11.22: Approach (b): Details of the precision, recall, F1 scores and Confusion Matrices (CM) for each combination of classifier and dataset, where CM (tn, fp, fn, tp) refers to the confusion matrix with values (True Negative, False Positive, False Negative, True Positive).