Electronic Thesis and Dissertation Repository

4-23-2021 10:00 AM

In-clinic Functional Measurement and Analysis of Knee Osteoarthritis Patients Undergoing Total Knee Replacement

Riley A. Bloomfield, *The University of Western Ontario*

Supervisor: Teeter, Matthew G., *The University of Western Ontario*
Joint Supervisor: McIsaac, Kenneth A., *The University of Western Ontario*
A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Electrical and Computer Engineering
© Riley A. Bloomfield 2021

Follow this and additional works at: https://ir.lib.uwo.ca/etd

Part of the Biomedical Devices and Instrumentation Commons

Abstract

Prevalence of osteoarthritis is increasing as individuals are remaining active later in life. Since the knee is one of the most commonly affected joints and is involved in almost all daily activities, functional impairment has a substantial impact on overall health. Despite this increase, there currently exists no disease modifying drugs or treatments. Mild cases are managed with physiotherapeutic exercises and common anti-inflammatories but surgical intervention is required for more severe disease progression.

Total knee replacement as a treatment for osteoarthritis is a highly successful surgery that is effective at restoring knee function and reducing pain but still requires further refinement. Over 70,000 of these surgeries are performed annually in Canada with 99% for the treatment of degenerative arthritis. Despite improvements to surgical technique and implant designs, studies report up to 20% of patients remain dissatisfied with their knee replacement up to the point of not undergoing the surgery again if it were an option. A singular cause for this dissatisfaction has not been pinpointed but strong influencers are pain, low functional improvement, and unmet expectations.

Early detection of functional problems permits further intervention through targeted physiotherapy or additional surgeries before problems escalate and cause patient dissatisfaction or implant revision. Current methods of patient evaluation rely on self-reported measures, which suffer from ceiling and floor effects often masking inter-patient differences. These measures are also influenced from patient expectations and what a patient reports they "can" do, is not always representative of their true functional ability.

Wearable sensors permit objective functional measurement of the knee as a supplement to patient-reported measures. Instrumented performance tests can measure patient function and compare to similar recoveries to highlight deficiencies or positive recovery traits. This

thesis outlines the development of such a wearable system for in-clinic measurement and the extraction of functional parameters to predict future outcomes and give surgeons the earliest indications for intervention. This information can also help surgeons realistically adjust patient expectations for recovery, even before undergoing surgery. It is expected that these individualized assessments to set expectations before surgical intervention will help address the persistently high patient dissatisfaction.

Keywords: Total knee replacement, Wearable sensors, Machine learning, Patient outcomes, Functional testing

Summary

Total knee replacement is a commonly performed surgery where the knee is replaced by an artificial joint. In Canada, 99% of these surgeries are performed as a treatment to knee osteoarthritis (OA): a disease that causes breakdown of cartilage, pain, and loss of joint function. Although exercise and medications are often prescribed for mild cases, this currently remains the only treatment available for late-stage knee OA.

Unfortunately given total joint replacement is the only solution to restore function and improve the quality of life of affected patients, up to one in five patients self-reports they are dissatisfied with their surgery and many indicating they would not undergo it again if it were an option. A variety of reasons for this dissatisfaction have been uncovered such as persisting pain, low joint function improvement, and expectations of how a new knee would perform not being met.

When undergoing knee replacement, evaluation tools to measure patients and track the joint recovery are limited. Complex equipment is not accessible for many clinics and physical space restrictions limit equipment that can be used. Questionnaires which instruct patients to self-report their abilities, pain, expectations, and satisfaction are most commonly used to get patient feedback. Unfortunately these can be unreliable when patients cannot distinguish subtle changes in their own health. Furthermore, perceived impairments could be caused by other joints (such as the hip) or conditions external to the knee which may unfairly influence answers.

The work presented in this thesis focuses on the development of a knee measurement tool that can be used in the clinic by patients and surgeons to get instant feedback on knee function. These measurements will be free of patient or surgeon bias, which permits results to be compared to previous visits or other patients. Later work describes the use of these measurements

to predict how much function a patient may regain even before their surgery, which can help the surgeon set realistic expectations for the patient, which in turn, is expected to help manage patient satisfaction.

Co-Authorship Statement

The following thesis contains material that has been published or submitted for publication. As the first author of these research articles, I have significantly contributed to the design, development, analysis, and compilation of all included published work. My supervisors Drs. Kenneth McIsaac and Matthew Teeter supervised my progress throughout the completion of all work and provided support for analysis and data interpretation. They were also responsible for final editing and review before submission of reports.

Chapter 4 includes work that was published in an original research article titled "Proposal and validation of a knee measurement system for patients with osteoarthritis" published in IEEE Transactions on Biomedical Engineering in 2018. This article was co-authored by Riley Bloomfield, Megan Fennema, Kenneth McIsaac, and Matthew Teeter. As the first author I was the main contributor to the manuscript preparation and data analysis presented. The robotic phantom testing and data collection was performed jointly with myself and the second author, Megan Fennema. Dr. McIsaac provided technical guidance during the development of the wearable sensor system while Dr. Teeter provided significant help editing the final article as well as substantial direction during the design of experiments.

Chapter 5 includes work that has been submitted for publication as an original research article titled "Wearable sensor instrumented timed-up-and-got test segmentation and metric extraction for knee replacement outcome assessment". This was submitted in 2021 and is under review at the time of writing. This manuscript was co-authored by Riley Bloomfield, Kenneth McIsaac, and Matthew Teeter. As the primary author I have substantially contributed to all aspects of preparing this manuscript. Dr. McIsaac provided technical guidance while Dr. Teeter contributed to editing the final submission. Data collection for the subjects under examination in this chapter was provided by Megan Fennema, who used the wearable sensor system detailed

in Chapter 4 to instrument subjects in the clinic during their clinical appointments.

Chapter 6 includes work from two original research articles titled "Machine learning groups patients by early functional improvement likelihood based on wearable sensor instrumented preoperative timed-up-and-go tests" and "Machine learning and wearable sensors at preoperative assessments: Functional recovery prediction to set realistic expectations for knee replacements" that have been published in Journal of Arthroplasty (2019) and Medical Engineering and Physics (2021) respectively. These articles were co-authored by Riley Bloomfield, Jordan Broberg, Harley Williams, Brent Lanting, Kenneth McIsaac, and Matthew Teeter. As the first author of both works, I substantially contributed to the design of experiments, analysis of data, formulation of results, and writing the articles. Jordan Broberg and Harley Williams performed the data collection for subjects analyzed in both works by instrumenting them using the wearable sensor system described in Chapter 4 at their clinical appointments. Dr. Lanting provided assistance to study design and clinical input. Dr. McIsaac provided technical support through the machine learning explored in the articles and Dr. Teeter provided substantial input to editing both articles as well as organizing data collection and providing study direction.

Chapter 7 includes work from an original research article titled "A convolutional neural network approach to classifying activities using knee instrumented wearable sensors" that has been published in IEEE Sensors Journal in 2020. This original research was co-authored by Riley Bloomfield, Matthew Teeter, and Kenneth McIsaac. As the primary author of this article I have substantially contributed to all aspects of study design, data collection, analysis, and presentation of the results. Dr. Teeter provided assistance to obtain institutional ethics approval to instrument subjects and helped with editing the article. Dr. McIsaac provided technical assistance to develop and test presented algorithms.

Acknowledgements

Contents

List of Figures

List of Tables

Nomenclature and Acronyms

Acronyms

ANN   Artifical Neural Network

BMI   Body Mass Index

CNN   Convolutional Neural Network

DCM   Direction Cosine Matrix

DT   Decision Tree

LHSC London Health Sciences Centre

LOS   Length of stay

MARG Magnetic, Angular Rate, and Gravity

NB   Naïve Bayes

OA   Osteoarthritis

OARSI Osteoarthritis Research Society International

PROM Patient reported outcome measure

ReLu   Rectified Linear

RF   Random Forest

SVM   Support Vector Machine

TKA   Total knee arthroplasty

TKR   Total knee replacement

TUG   Timed-up-and-go

UUID Universally Unique Identifier

Units

$(\deg/s)$ Degrees per second

deg Degrees

kg/m^2 Kilograms per metre squared

Chapter 1

Introduction

Total knee replacement (TKR) is a highly successful elective procedure consisting of replacing the articulating surfaces of the knee with artificial components. In Canada, over 70,000 knees are replaced annually with 99% of these surgeries performed as a treatment for end-stage knee osteoarthritis (OA) [1]. General goals of this surgery include reducing pain, restoring joint function, and increasing joint stability to improve patient quality of life.

Annual knee replacement surgeries have increased by 17% in the last five years, rising alongside increasing rates of comorbidities such as obesity. The number of older individuals remaining active later in life is also increasing, straining joints with more vigorous activities [1]. A growing number of joint replacement surgeries is placing a burden on our health care systems, contributing to long wait times during which the patient's health often declines further due to decreased mobility and limited daily activity. Despite being a comparably successful surgery, there are several important motivators for additional research and investigation:

- There is a high 7% rate for knee revision. This process involves the removal of an existing implant and a second, more complicated surgery is performed to insert another. Healthcare costs of a revision surgery alone (not including increased recovery support) are 80% higher than a primary implant and patients stay in the hospital twice as long on average (8.9 days vs. 4.2 days) [1]. Additional days in the hospital are associated with increased risk of infection and other complications.

- Knee OA affects a wide age demographic including both aged and younger populations.

In the United States there are 14 million people suffering from symptomatic knee OA; six million who are under age 65, and two million under age 45 [2]. Younger patients tend to be more active and are more likely to be in the workforce and to have more demanding joint requirements for their livelihood. A 65 year old patient has a 7% lifetime risk of requiring a knee replacement revision but this substantially increases for younger patients who are more commonly active and require increased implant longevity [3]. Implants must be continuously evaluated to ensure they are performing as expected for extended periods in all patient groups.

- Since this elective, highly successful surgery is the only option to improve the quality-of-life of afflicted patients, it is especially troubling that a 20% self-reported surgical dissatisfaction rate persists in many studies and across jurisdictions, with many patients reporting they would not undergo the surgery again given it were an option [4–8].

After decades of knee replacement research and optimization, this high patient dissatisfaction rate persists across many jurisdictions. Furthermore, dissatisfied patients draw healthcare resources away from other individuals waiting for surgery. They may seek additional appointments or calls to their surgeon for inquiries and spend longer at their appointments getting their concerns addressed and although a vague measure, satisfaction is of increasing interest to health payers [9, 10]. Satisfaction has been linked to preoperative function and functional improvement after surgery with more advanced OA patients being more satisfied with the procedure, however, the greatest risk factor for dissatisfaction is unmet patient expectations [4, 5, 11–14]. Bourne et al. found that low postoperative outcome scores, high preoperative pain, or postoperative complications requiring rehospitalization were all associated with a 2-3× greater risk of dissatisfaction whereas patients with unmet expectations were at 11× greater risk [4]. Connor-Spady et al. have identified that expectations concerning joint function and physical activity are least often met, suggesting many patients expect more functional recovery than is reasonably likely, and patients were significantly more satisfied when more of their expectations were met [15]. Managing and appropriately adjusting patient expectations before surgery, with emphasis on functional improvement, has been identified as one of the most influential methods of improving satisfaction post-TKR but this has remained impractical with traditional methods of

patient evaluation [6, 15].

1.1 Motivation

Due to time constraints on patient clinical exam visits and the complexity of available evaluation equipment, in-clinic patient evaluation has shifted to self-reported measures. Despite care providers becoming increasingly pressured to demonstrate benefits of TKR using tangable, objective measures, patient reported outcome measures (PROMs) have become increasingly common tools for patient evaluation [3, 16]. They are an important population-level assessment tool for investigating general best practices and have been considered the gold standard for assessing surgical outcomes [17–20]. Unfortunately these subjective, self-reported measures are not without limitations: they fluctuate with patient moods, are biased by current experiences and expectations, and self-reported functional ability has been shown to differ from objectively measured performance, highlighting the need for more granular and accurate measurement of function [21, 22]. They also experience floor and ceiling effects, which makes it difficult to distinguish subtle differences in patient status, and limits the effectiveness of individualized assessments [23, 24]. As a research tool, these patient-completed surveys are inefficient when administered in the traditional hard-copy medium. Long paper questionnaires are filled by patients and often are not entirely completed, rendering the partial data unusable. Even when surveys are filled correctly, responses must be logged, accumulated, and tallied which requires additional time by the researcher.

Tools for obtaining objective metrics to evaluate patient function exist such as motion capture and gait assessment labs. Current methods for measuring knee angles and usage such as instrumented gait analysis or camera-based measurements are expensive and must occur in testing environments that may not be easily accessible to patients. Since these traditional options require trained staff to operate, only limited data collection is feasible. Additionally, it has been indicated that a Hawthorne effect sometimes influences measurement accuracy when patients are being closely observed by an evaluator, which produces more ideal and less natural patient data than would normally be recorded if unsupervised [25, 26].

Continuous instrumentation of knee angles and other functional parameters without con-

straining measurements to a specific environment would allow acquisition of quantitative joint metrics during activity without complex and obtrusive instrumentation. Information provided could accurately determine how individual patients use their joint by monitoring angles over time, pre/post-surgery, and how this correlates to their satisfaction to identify key joint characteristics or functional activities that correlate to patient satisfaction using quantifiable metrics on a per-patient basis. With this knowledge, more individualized care can be provided to help promote satisfaction and identify any functional deficiencies along recovery.

1.2   General Problem Statement

Ideally, every patient should be functionally assessed before and after undergoing TKR at all their clinical appointments to measure baseline functional performance and check for subsequent functional improvement following surgery. This assessment provides surgeons and clinicians with a measure of patient function but also provides patients with information concerning their improvement, which can be used to adjust their personal expectations or change their recovery habits to give them the best chance at regaining the function they expect. Currently, this is impractical due to limited equipment availability and clinical appointment times. The current standard of care relies on patients to communicate any perceived or potential problems. Patients may not notice any subtle decline in function, and it is even less likely that these observations would propagate through self-reported measures. The current standard of care relies on surgeons and clinicians to determine if there are any problems at checkups, but it may be possible to go further and not only determine problems, but identify positive traits indicative of positive outcomes further through recovery.

With increasing surgeries being performed in increasingly busy clinics, surgeons will not have more time available to perform additional observation and measurement of patients with the current standard of care. This work seeks to provide a portable method of instrumentation that can be used by surgeons and other healthcare staff to obtain granular, objective measurements and clinically relevant information that can help better provide individualized care plans and suggestions.

1.3 Research Objectives

The overall goal of this thesis is to investigate the utility of performing granular, objective functional measurement of patients undergoing total knee replacement. This was achieved through the following breakdown of objectives:

- Determine an appropriate sensor technology to acquire functional performance measurements of patients in restricted environments, such as orthopaedic clinics and examination rooms where space is limited and complex equipment is not practical.

- Develop and validate a software system using an algorithm to extract granular joint-specific metrics to objectively measure knee performance from raw sensor outputs.

- Investigate the utility of repeatable functional tests for patients to complete in a clinical setting using minimal additional equipment to measure joint performance across multiple instrumentation sessions and patients.

- Investigate applications of derived joint performance metrics to predict patient functional recovery outcomes following their surgery using information from before their operation.

- Investigate the capability of performing activity recognition using only a knee measurement system for the extension of this work to less rigid functional testing that make take place in varying unscripted environments.

1.4 Scope

The scope of this work includes the design considerations and development of an appropriate instrumentation system to measure patients' knee function in the clinic. Algorithms were developed to parse recorded sessions of activity to extract both clinically relevant parameters and abstract metrics to analyze patient motion. Further research into the assessment of patients using the developed system is demonstrated, primarily using machine learning, for the translation of this work into clinical practice. The work includes several novel proof-of-concept developments that are intended as a starting point for further patient analysis.

1.5 Thesis Structure and Contributions

The remainder of this thesis is divided into the following chapters:

Chapter 2   Background: provides relevant background in the areas of current joint replacement, current outcome assessment, wearable sensors, inertial sensors and raw data representations, functional testing, and machine learning.

Chapter 3   Literature review: summarizes the state of the art in knee instrumentation using wearable sensors, machine learning to predict functional recovery outcomes in joint replacement, and human activity recognition using wearable sensors.

Chapter 4   Wearable sensor system development: details the selection of appropriate wearable sensors to instrument patients in the clinic and the development of a software strategy to extract objective joint parameters from raw sensor data.

Chapter 5   Deriving performance test measures: presents the selection of an appropriate functional test for patients to perform while instrumented with the developed wearable system and the extraction of clinically relevant joint metrics to evaluate patient function.

Chapter 6   Predicting functional recovery: describes the application of unsupervised machine learning to discover parameters that influence patient functional recovery and the development and training of a classifier to predict short term functional recovery using only preoperative patient functional metrics.

Chapter 7   Activity recognition: presents an application of the developed wearable knee measurement system to recognize daily activities performed.

Chapter 8   Conclusions and future work: highlights the contributions presented in this thesis and proposes areas for future continuation of this work.

Chapter 2

Background

2.1 Introduction

This chapter will provide an overview into topics explored and developed in the remainder of the work. Topics concerning the clinical background of osteoarthritis and its treatments are covered before moving on to current standards and methods for patient evaluation. Material on existing methods of knee instrumentation will follow, leading into a background of relevant wearable technologies. Due to the use of extensive raw orientation measurement in the following chapters, a summary of orientation representations will be presented. Lastly, an outline of applicable machine learning techniques will be introduced with further details on specific models and methods presented where they are used in future chapters.

2.2 Knee Osteoarthritis

Osteoarthritis (OA) is the most common joint disease in the world and is one of the most frequent causes of pain, disability and loss of function [27]. It is a disease of the entire joint, caused by the degeneration of articular cartilage and the underlying bone. Traditionally OA was considered a result of joint "wear and tear" but occurs in joints with various workloads or traumas and is influenced by genetics [28]. The most commonly affected joints include knees and hips, with the incidence of knee OA being more common. It can be diagnosed clinically if patients report joint pain or impairments such as morning stiffness by palpating

7

for bony enlargement or audible indicators through joint motion [29]. OA is confirmed with radiographs mainly using the Kellgren and Lawrence score which is a grading based on the narrowing of the joint gap and other visual indications [30].

2.2.1 Prevalence, Risk Factors, and Impact

In 2010, the global prevalence of radiographically diagnosed symptomatic knee OA was nearly 4% and is one of the top 25 most globally prevalent sequelae [30, 31]. Developed countries have increased prevalence compared to developing and the incidence is expected to rise with an aging population that is remaining increasingly active later in life [30, 32, 33]. Prevalence of knee OA peaked around age 50 and was most prevalent in Asia Pacific region but lowest in South and Southeast Asia [32]. In the United States, there are over 14 million individuals living with symptomatic knee OA; 8 million that are under the age of 64 [2]. According to the Arthritis Society, OA affects nearly one in six Canadians and is expected to grow to one in four by 2035. Primary risk factors for developing knee OA are age, gender, obesity and genetics [30]. Lower extremities are crucial for completing day-to-day tasks and daily activities and impairments make it difficult to remain active. Loss of joint function and reduced activity levels increase the risk of cardiovascular disease, diabetes, and cancer [34, 35].

Functional impairment due to OA is impactful to our health system. In Canada alone, productivity costs due to work loss associated with OA are expected to reach $17.5B CAD by 2031 [36].

2.2.2 Treatments

Early stages of knee OA can be treated with physiotherapy or increased exercise since these have been shown to provide short-term functional improvement and pain reduction. It is suggested that land and water-based exercises, weight management, and strength training are among the most effective non-invasive treatments. Walking aids such as a cane are recommended for patients with single knee OA however the redistributed stress from the aid may cause discomfort in other affected joints [37]. Strengthening the muscles surrounding the joint can increase support, stability, and improve joint function but does not prevent further worsen-

ing of the articulating surfaces.

As the disease progresses and degeneration continues to occur, these methods may fail to improve patient functional performance and typically medications are prescribed to alleviate pain. This is a non-sustainable solution as the disease will continue to progress and further loss of function is inevitable: there are currently no disease-modifying drugs available to prevent or reverse the progression of OA. Although the advancement of biological methods to enhance cartilage reproduction seem promising, it may be many years before they are widely available and currently, surgical intervention is most common solution [38].

Injections and regenerative medicines have been suggested to offer pain relief and have the potential to treat knee OA however their effectiveness and therapy progression limit their suitability as a long-term solution. Corticosteroids have been injected into the affected joint for decades as method of prolonging surgical intervention but their usefulness concerning slowing or stopping disease progression is still debated [36].

2.3 Total Knee Replacement

Total knee replacement (TKR) is a highly successful surgery performed as a solution for end-stage knee OA. It is an elective, quality-of-life improving surgery with the goal of reducing joint pain and improving function, permitting patients to complete daily activities and stay mobile and active.

In the United States, more than 700,000 knees are replaced are annually and conservative estimates show this number increasing despite economic downturn [3]. In Canada, with our smaller population, approximately 70,000 replacements are performed annually which has increased 17% over the last five years and is expected to continue to grow. The mean acute length of stay (LOS) in hospital was three days with only 0.4% of patients returning home the day of their surgery [1]. The average age of patients undergoing this surgery is 65 years with progressively more surgeries being performed on younger patients [3]. Almost all knee replacements in Canada are performed as an end-stage treatment for OA (99%) with few others to treat injuries or joint trauma [1].

The knee consists of bone structures, cartilage, ligaments and a synovial membrane which

contains synovial fluid that is used to lubricate the joint. During a TKR, the articulating surfaces of the knee are replaced with metal or ceramic prostheses and these are separated by a polyethylene liner.

The surgical process generally follows bone preparation, insertion of the artificial components, then the plastic spacer though replacements can be performed differently at the surgeon's discretion. A variety of implant designs can be used with variable-length stems (both cemented and cement-less) inserted with standard or patient-specific tools. When deciding surgery configurations, reproducing the best outcomes across patient populations and restoring functional ability are highly influential as reported by orthopaedic surgeons [39]. Once an implant has been inserted, there are further choices concerning balancing the knee; a process to ensure the joint will function without excessive or unbalanced implant contact during motion. Some surgeons believe a joint gap-balancing technique will provide a more stable joint while others insist a measured resection is more effective (components are aligned based on bone landmarks) [40].

2.4 Patient Evaluation

Since TKR is an elective, quality-of-life improving surgery, it remains practical to collect patient reported outcome measures (PROMs) for surgical evaluation. However, due to the limitations mentioned in Chapter 1, self-reported measures alone are not sufficient to properly assess patients and should be supplemented with objective measurements [24]. Patient specific evaluation is motivated by studies indicating a large difference in outcomes measured by a clinician over patient-specific measures [4].

PROMs are widely used tools for obtaining qualitative feedback pre- and post-surgery. Measurements of health are obtained from patients' perspectives through surveys and questionnaires. A common example for hips and knees is the WOMAC patient questionnaire (Western Ontario and McMaster Universities Arthritis Index) [41]. This assessment method has been proven valid when administered over the phone, electronically, and through hard-copy in person, which greatly facilitates information retrieval [42]. PROMs have been praised for their ability to capture quality of life metrics through patients' perspectives but are not without

faults. Minor changes in health may not be noted in self-evaluations while small improvements to function or pain may be over-amplified, resulting in ceiling effects [43]. WOMAC may also be less effective for patient measurement shortly following surgery and may be more effective at further time points [44]. Activity levels and functional performance have been shown to differ when self-reported compared to measurements from instrumented sessions, especially in overweight and obese patients [21, 22]. Similarly, self-evaluated improvement in performance of activities does not correlate with objectively measured function in OA patients following joint replacement which emphasizes the need for quantitative instrumentation [22]. While useful for registering patient qualitative data, this evaluation method fails to provide quantitative joint measurements or objective measurement of the joint free of patient bias.

2.5   Knee Instrumentation

More objective methods of patient evaluation currently exist although they are not commonly deployed for practical reasons. Viable options include complex camera-based motion capture systems relying on depth cameras, reflective markers, or recognized object tracking. Unfortunately it is unrealistic to expect the standard of care to include use of motion capture laboratories for all patients at multiple time points along their surgery and recovery.

Time required to transport patients to and from these types of facilities can be extensive, if the technology is accessible at all. Patients may also become fatigued before the instrumentation sessions which will influence their results. In addition to the cost of acquiring the mo-cap system, an operator must be employed to run the capture software during patient testing. Lastly, the field of view is often limited in all but the best of systems so that any functional testing will have to be constrained to the confines of the testing setup. To avoid the dedicated laboratory constraint, research has investigated low cost depth cameras for tracking lower extremities and removes the dependence on reflective markers, mainly for gait cycle detection [25]. Similar systems have also been investigated involving normal RGB video cameras as well with the same limitations; there is a limited field of view for instrumentation and patients cannot venture far from the camera's range. Additionally there can not be any obstructions interfering with camera contact so testing in a home or ambulatory setting with obstacles or

uncalibrated cameras is not an option. Additionally with more cumbersome instrumentation, it has been indicated that a Hawthorne effect sometimes influences measurement accuracy when patients are being closely observed by an evaluator or if they can feel instrumentation equipment, which produces more ideal and less natural patient data than would normally be recorded if unsupervised [25, 26].

A more appropriate system to instrument patients to observe their knee joint function must be portable, unobtrusive, low-cost, and easily deployable with few constaints on the environment. Continuous instrumentation of knee angles without constraining measurements to a specific environment would allow acquisition of quantitative joint metrics during usage without complex and obtrusive instrumentation. Information provided could accurately determine how individual patients use their joint by monitoring angles over time, potentially pre/post-surgery, and how this correlates to their satisfaction. This could identify key joint characteristics or functional activities that correlate to patient satisfaction using quantifiable metrics on a per-patient basis. Several technologies have been explored as portable instrumentation solutions to measure patient function in varying settings.

2.6 Wearable Sensors

Wearable sensors are currently used in many industries including film animation, video game development, motion capture and medical instrumentation. Wearable systems can be small-sized and provide a viable method of instrumenting patients to assess surgical outcomes and detect movement disorders without constraining their environment [45]. A variety of wearable sensor types exist and can be split into two main categories: flexible materials and inertial units.

Flexible Materials

If tight fitting fabric can be worn by patients, flexible sensing materials provide an excellent option for obtaining joint angles. Multiple sensing fibres can be placed around body joints to accurately measure joint angles in multiple dimensions [46, 47]. These small fibres are lightweight but are limited by the physical length of the fibre; they cannot be separated. To

measure joint angles, these devices must be mounted proximal to the affected joint and held in place, often with tight fitting materials or adhesives to prevent movement as the fibres stretch. Measurement accuracy could vary with the alignment of these devices in the joint coordinate frame. This could cause discomfort for post-operative patients with joint swelling or pain. Devices placed near or surrounding an incision could increase the risk of infection, a devastating complication that must be avoided. Similarly, digital goniometers have been used to determine joint angles with similar usage constraints [48,49]. Goniometers produce low-noise and precise measurements but can be obtrusive for patients when combined to measure multiple degrees of freedom and their accuracy is dependent on correct sensor placement [45].

Inertial Units

An inertial measurement unit (IMU) is a micro-electromechanical system (MEMS) comprised of a gyroscope and an accelerometer. Most modern IMUs contain tri-axial components, allowing operation in three dimensions. Data extracted from these devices can be raw components readings or an estimate of the devices orientation in space computed using a combination of the raw component readings. Orientations are expressed about a coordinate frame and when discussing coordinate frames, it is also important to distinguish between a local frame and a global one. A local frame can be assigned to an object rather arbitrarily however, it is important that the frame remain fixed to the object and that it rotates along with changes in the moving body's rotation. A global frame will remain fixed about a world axis such that changes in rotation to bodies moving in the world do not alter the world frame. This represents a fixed global frame that objects can be measured relative to. The assignment of a global frame can also be arbitrary but this frame must be considered a fixed reference, relative to any other rotating bodies. A depiction of a local frame relative to a fixed global world frame can be seen in Figure 2.1.

Gyroscopes: A modern gyroscope measures a change in angular velocity, measured in (deg $/s$). A tri-axial component will measure this change of angular velocity about each of three orthogonal axes independently (Figure 2.2). Given a fixed amount of time t, a previous orientation estimation of $\mathbf{O_0}$, and a tri-axial gyroscope reading of $\mathbf{R_{gyro}}$ in $\frac{deg}{t}$, the current orientation $\mathbf{O_1}$ can be estimated as:
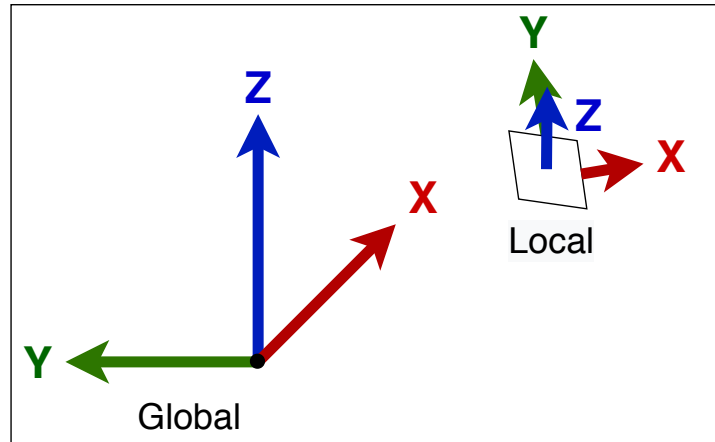
Figure 2.1: A global reference frame is shown on the left, which will not change due to rotations in the local body on the right. The local frame on the moving body will remain fixed to the object and change with rotations on the body.

$$\mathbf{O_1} = \mathbf{O_0} + (\mathbf{R_{gyro}} \times t) \tag{2.1}$$

Accelerometers: An accelerometer measures external accelerations acting on the device. If the sensor is at rest, it would be expected that the only acceleration detected would be gravity at 9.8 m/s^2. At rest, there are no other external accelerations however, during motion, many external forces will be acting on the device and it is less convenient to obtain the direction of gravity to aid in obtaining an orientation estimate.

Magnetometer: Magnetic angular rate and gyroscope (MARG) sensors are similar to IMUs, but incorporate magnetometers. The purpose of this additional sensing component is to locate magnetic north using Earth's magnetic field. An additional absolute reference can be used to adjust gyroscope and accelerometer readings through extended motions affected by abrupt changes in motion and gyroscope drift.

Sensor Fusion

Sensor fusion algorithms take raw gyroscope (deg $/s$) and accelerometer (m/s^2) readings and output an orientation estimation with respect to a calibrated coordinate system. This calibrated coordinate frame is referred to as the *global*, or *fixed* frame because it does not change

(a) A right-handed tri-axial coordinate frame.

(b) A left-handed tri-axial coordinate frame.
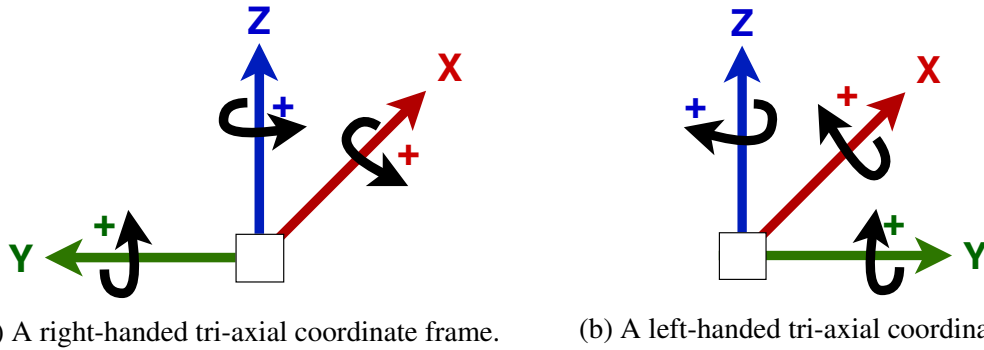
Figure 2.2: Coordinate frames for both left and right handed coordinate systems. A positive rotation is indicated by the arrows for each figure. In a right-handed frame, a counterclockwise rotation about an axis is positive while a clockwise rotation is positive in a left-handed frame.

with the moving body being measured. If measurements were not with respect to this global frame, it would be impossible to compute changes in body orientation, or compare the orientations of multiple local bodies, unless they are measured with respect to the same global frame (Figure 2.1).

Often the raw values of sensor fusion components (gyroscope and accelerometer) can be queried independently to implement custom algorithms or set a transformed local coordinate frame, such as the joint convention when coupled to a body segment. Generally, IMUs produce drift over extended use if they do not return to rest allowing a true reading of gravity to be obtained. In this scenario, a magnetometer can be used to find magnetic north. When IMUs incorporate a compass, they are also referred to as magnetic, angular rate, and gravity (MARG) sensors. This additional reading is used to obtain the direction of Earth's magnetic field to help orient the sensor when external accelerations contribute to false accelerometer-based corrections. This hardware can generally obtain more accurate orientation estimates, but may be influenced by external magnetic interference [50, 51]. MARG sensors are also referred to as attitude and heading reference systems (AHRS) as they are used in aircraft and autonomous flight applications as a replacement for traditional mechanical attitude and heading devices [52].
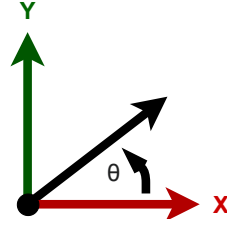
Figure 2.3: A rotation θ about the vertical axis (protruding from image) orthogonal to the *xy* plane shown.

2.6.1 Orientation Representations

Orientation estimations with respect to a reference point, or fixed global frame, can be expressed in several ways: most commonly as rotational matrices; cardan (Tait-Bryan) and Euler angles; and quaternions.

Rotation Matrices

A three-dimensional rotation matrix is a 3x3 matrix used to describe the rotation of a point in a three-dimensional Cartesian coordinate system. In a right-handed system (most common), a counterclockwise rotation about an axis is considered positive.

A more simple 2D rotation of a point θ in two dimensions on an *xy* plane (Figure 2.3) can be expressed as:

$$Rot_{xy}(\theta) = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$$

where θ is a rotation about the axis orthogonal to this *xy* plane (Figure 2.3).

Considering a three dimensional frame (Figure 2.2a), this same rotation is expressed as:

$$Rot_z(\theta) = \begin{pmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{2.2}$$

where the matrix has been expanded to a 3x3 form, but with no additional rotation about either the *x* or *y* axes. By observing different axes of a 3D frame, it can be seen that independent rotations about either the *x* or *y* axes are expressed as:

$$Rot_x(\theta) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta) & -\sin(\theta) \\ 0 & \sin(\theta) & \cos(\theta) \end{pmatrix} \tag{2.3}$$

and

$$Rot_y(\theta) = \begin{pmatrix} \cos(\theta) & 0 & \sin(\theta) \\ 0 & 1 & 0 \\ -\sin(\theta) & 0 & \cos(\theta) \end{pmatrix} \tag{2.4}$$

Since a rotation or the change in orientation of a body relative to a global reference frame is rarely only a change about one axis, rotations about multiple axes must be considered. Any change in orientation of a body about multiple axes can be described as a series of three independent rotations about independent axes, each compounding on the previous. It must be noted that the order each of these three rotations is performed is relevant since matrix multiplication is not commutative, and will result in a different final body orientation if performed in an alternate order. Therefore, when orientation is described using a combination of rotations, the rotation sequence must also be known. Given that before any rotations are performed on a moving body, the local frame of the body can be considered aligned with the global frame. If each of the three rotations in a sequence are performed about the global world frame, this change is described as an extrinsic rotation.

Alternatively in an intrinsic sequence, each independent rotation is performed on the transformed local body frame that was altered in the previous rotation stage. As long as the three rotation magnitudes are equivalent, an intrinsic sequence results in the same change in orientation as an extrinsic sequence when the order of rotations are reversed. For example, an extrinsic sequence $Z_\theta Y_\psi X_\rho$ is equal to an intrinsic rotation sequence of $X_\rho Y_\psi Z_\theta$.

Since rotations are performed in three stages about each body frame axis independently, and sequential rotations about the same axis are equivalent to a single rotation of a different magnitude

$$Rot_x(\theta_1) + Rot_x(\theta_2) = Rot_x(\theta_1 + \theta_2),$$

there are $3 \times 2 \times 2 = 12$ possible rotation sequences. As an example, the sequence XYZ

describes a rotation about x first, a second rotation about y, and a final rotation about z. When computing the combined rotation of this sequence from each independent rotation, rotations are postmultiplied for extrinsic rotations (rotations about the global frame), and premultiplied for intrinsic rotations (compounding rotations about the transformed local frame).

A change in orientation of a local body $Rot_{xyz}(\phi)$ rotating by ρ about the global x axis, then θ about the global y axis, and finally ψ about the global z axis (extrinsic) is shown as:

$$Rot_{xyz} = Rot_x(\rho)\, Rot_{y'}(\theta)\, Rot_{z''}(\psi)$$

$$= \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\rho) & -\sin(\rho) \\ 0 & \sin(\rho) & \cos(\rho) \end{pmatrix} \begin{pmatrix} \cos(\theta) & 0 & \sin(\theta) \\ 0 & 1 & 0 \\ -\sin(\theta) & 0 & \cos(\theta) \end{pmatrix} \begin{pmatrix} \cos(\psi) & -\sin(\psi) & 0 \\ \sin(\psi) & \cos(\psi) & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} \cos(\psi)\cos(\theta) & -\cos(\theta)\sin(\psi) & \sin(\theta) \\ \cos(\rho)\sin(\psi) + \cos(\psi)\sin(\rho)\sin(\theta) & \cos(\psi)\cos(\rho) - \sin(\psi)\sin(\rho)\sin(\theta) & -\cos(\theta)\sin(\rho) \\ \sin(\psi)\sin(\rho) - \cos(\psi)\cos(\rho)\sin(\theta) & \cos(\psi)\sin(\rho) + \cos(\rho)\sin(\psi)\sin(\theta) & \cos(\rho)\cos(\theta) \end{pmatrix}$$

$$(2.5)$$

Rotation matrices contain nine elements to represent a change in orientation in 3D space. It can be seen through Equation (2.5) that some of these elements are dependent. Only three parameters are required for a minimal representation of these rotations.

Cardan Angles

When a rotation is performed about **all three** axes of a body frame, the minimal representation of the rotation sequence is known as a set of cardan angles or also as *Bryan-Tait* angles. These possible rotation sequences are *XYZ*, *XZY*, *YXZ*, *YZX*, *ZXY*, and *ZYX*. By first imposing a general rotation matrix structure for a 3x3 matrix with i rows and j columns to indicate each element:

$$R = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix} \tag{2.6}$$

three cardan angles can be solved as an inverse solution to Equation (2.6). Given the result of Equation (2.5), the angles ρ, θ, and ψ can be found as:

$$\rho = atan2\left(\frac{r_{23}}{r_{33}}\right)$$
$$= atan2\left(\frac{-\cos(\theta)\sin(\rho)}{\cos(\rho)\cos(\theta)}\right)$$
$$= -atan2\left(\frac{\sin(\rho)}{\cos(\rho)}\right), \rho \neq \frac{\pi}{2}, \frac{-\pi}{2}$$

$$\theta = asin(r_{13})$$
$$= asin(\sin(\theta)), -\frac{\pi}{2} \leq \theta \leq \frac{\pi}{2}$$

$$\psi = atan2\left(\frac{r_{12}}{r_{11}}\right)$$
$$= atan2\left(-\frac{\sin(\psi)}{\cos(\psi)}\right)$$
$$= -atan2\left(\frac{\sin(\rho)}{\cos(\rho)}\right), \rho \neq \frac{\pi}{2}, \frac{-\pi}{2}$$

One common sequence of cardan angles is the *roll − pitch − yaw*, or *ZYX* set of rotations performed as an intrinsic sequence (Figure 2.4). This is a common representation in the aeronautical field and can be used to explain the changes in attitude of an aircraft.

The *roll − pitch − yaw* angles of a body rotating with respect to a fixed frame can also be computed by combining basic rotation matrices. The change in orientation of a local body $Rot_{xyz}(\phi)$ rotating by ρ about the local x axis, then θ about the body's transformed local y axis,
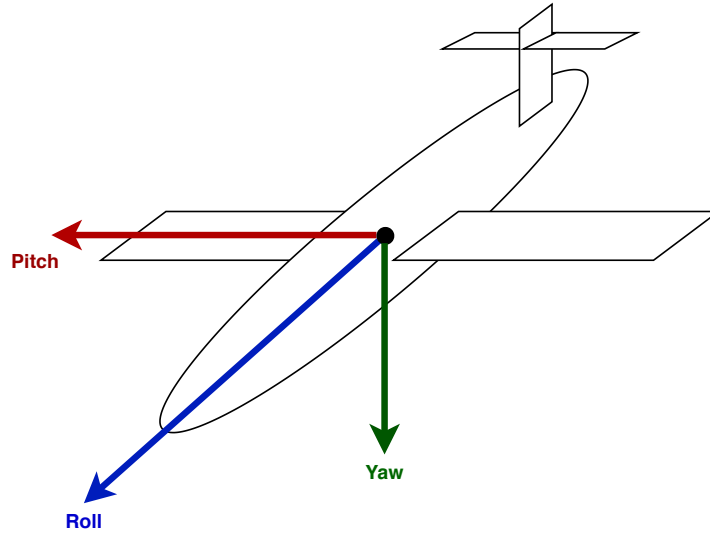
Figure 2.4: A *roll − pitch − yaw* rotation frame.

and finally ψ about the body's twice transformed z axis (intrinsic) is shown as:

$$Rot_{zyx} = R_z(\rho)R_y(\theta)R_x(\psi)$$

$$= \begin{pmatrix} \cos(\psi) & -\sin(\psi) & 0 \\ \sin(\psi) & \cos(\psi) & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \cos(\theta) & 0 & \sin(\theta) \\ 0 & 1 & 0 \\ -\sin(\theta) & 0 & \cos(\theta) \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\rho) & -\sin(\rho) \\ 0 & \sin(\rho) & \cos(\rho) \end{pmatrix}$$

$$= \begin{pmatrix} \cos(\psi)\cos(\theta) & \cos(\psi)\sin(\rho)\sin(\theta) - \cos(\rho)\sin(\psi) & \sin(\psi)\sin(\rho) + \cos(\psi)\cos(\rho)\sin(\theta) \\ \cos(\theta)\sin(\psi) & \cos(\psi)\cos(\rho) + \sin(\psi)\sin(\rho)\sin(\theta) & \cos(\rho)\sin(\psi)\sin(\theta) - \cos(\psi)\sin(\rho) \\ -\sin(\theta) & \cos(\theta)\sin(\rho) & \cos(\rho)\cos(\theta) \end{pmatrix}$$

$$\tag{2.7}$$

Solving the inverse solution of the format shown in Equation (2.6), the three independent angles ρ, θ, and ψ can be found as:

$$\rho = \text{atan2}\,(r_{21}, r_{11})$$

$$\theta = \text{atan2}\left(-r_{31}, \ \sqrt{r_{32}^2 + r_{33}^2}\right)$$

$$\psi = \text{atan2}\,(-r_{32}, -r_{33}) \tag{2.8}$$

Euler Angles

Proper Euler angles or Euler sequences require the first and last axis of rotation to be the same. An easily seen benefit to this is that extrinsic and intrinsic rotation sequences are both formed using the same combination of basic rotations. It is important to note that this does not necessarily mean intrinsic and extrinsic Euler rotations are equivalent since different magnitudes of rotations can be performed even if the axis of rotation is the same. The possible Euler angle sequences are *XYX*, *XZX*, *YXY*, *YZY*, *ZXZ*, and *ZYZ*. Rotations can be computed similarly to methods shown in Equation (2.5) or Equation (2.7).

Direction Cosine Matrix

A direction cosine matrix (DCM) is a rotation matrix that describes the rotation from one three-dimensional frame to another as a single rotation, partially rotating about any of the axes instead of three independent rotations about each axis of the body's local frame. This can also be used to describe the rotation from a global frame to a rigid body's local frame to indicate its orientation relative to the global frame.

The columns of a DCM represent unit vectors of an axis of the moving body frame projected on the global reference frame and each of the three rows represents the x, y and z dimensions.

$$Rot_{DCM} = \begin{pmatrix} c_{1x} & c_{2x} & c_{3x} \\ c_{1y} & c_{2y} & c_{3y} \\ c_{1z} & c_{2z} & c_{3z} \end{pmatrix} \tag{2.9}$$
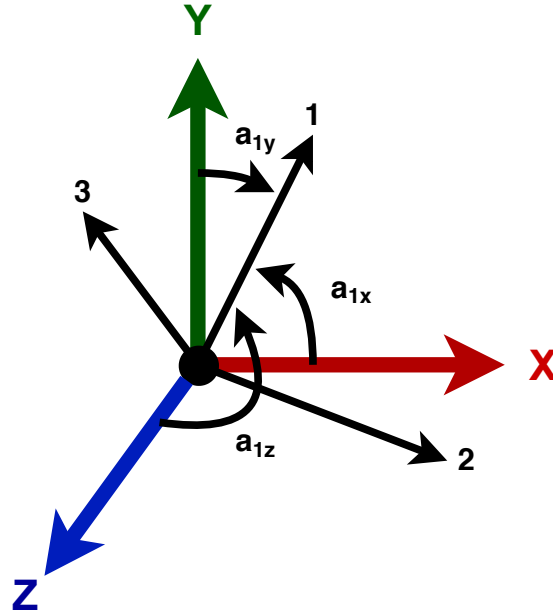
Figure 2.5: Given a global reference frame with axes *xyz* and local frame with axes 123, the elements c_{1x}, c_{1y}, c_{1z} of the direction cosine matrix shown in Equation (2.9) are found to be the cosine of angles a_{1x}, a_{1y} and a_{1z} respectfully. Elements of the remaining two columns are cosines of angles between axes 2 and 3 and the *x*, *y* and *z* axes of the global frame.

Quaternions

A quaternion is a minimal and unique representation of a body's orientation in three dimensional space. It is commonly used in graphics and computer applications because of its avoidance to singular configurations and efficiency performing multiple rotations as linear combinations. A quaternion consists of only four numeric components, making it computationally and memory efficient.

A quaternion **q** representing a rotation A with respect to a global coordinate system G is defined as: ${}_{A}^{G}\boldsymbol{q} = [q_0, q_1, q_2, q_3] = a + b\hat{i} + c\hat{j} + d\hat{k}$, where a, b, c, d are real scalar components and $\hat{i}, \hat{j}, \hat{k}$ are orthogonal quaternion units. Each of the four numeric components q_0, q_1, q_2, q_3 are in the range $[-1, 1]$.

The conjugate of a quaternion **q** describing a rotation A with respect to G is equivalent to its inverse and is defined as:

$$\,{}_{A}^{G}\boldsymbol{q}^* = {}_{G}^{A}\boldsymbol{q} = [q_0, -q_1, -q_2, -q_3] \tag{2.10}$$

We can define quaternion multiplication as a combination of two rotations **q** and **r** as:

$$q \otimes r = \begin{bmatrix} (r_0 q_0 - r_1 q_1 - r_2 q_2 - r_3 q_3) \\ (r_0 q_1 + r_1 q_0 - r_2 q_3 + r_3 q_2) \\ (r_0 q_2 + r_1 q_3 + r_2 q_0 - r_3 q_1) \\ (r_0 q_3 - r_1 q_2 + r_2 q_1 + r_3 q_0) \end{bmatrix}^T \tag{2.11}$$

This combination of two quaternion rotations shown in (2.11) is non-commutative. A local rotation **r** can be applied to a quaternion **q** to rotate it around its local frame as follows

$$q_{local} = r \otimes q \tag{2.12}$$

while **q** is rotated globally by a rotation **r** as

$$q_{global} = q \otimes r \tag{2.13}$$

The rotational difference between two quaternions $_G^A q$ and $_G^B q$ can be found as:

$$_A^B q = {_G^A q}^* \otimes {_G^B q} = {_A^G q} \otimes {_G^B q} \tag{2.14}$$

For the current work, unit quaternions will be used exclusively which have been normalized as:

$$\mathbf{q} = \frac{q}{\|q\|} = \frac{q_0, \ q_1, \ q_2, \ q_3}{\sqrt{q_0^2 + q_1^2 + q_2^2 + q_3^2}}$$

2.6.2 Anatomical Knee Angles

Often times the knee is considered to move as a hinge with rotation restricted about a single axis. In reality, the knee rotates in three axes during motion. As shown in Section 2.6.1, there are several ways to note rotations in three dimensions. The joint axes of the knee are commonly described using methods presented by *Grood and Suntay* where the femoral (thigh) and tibial (shin) segments are first located using their local coordinate frames [53].
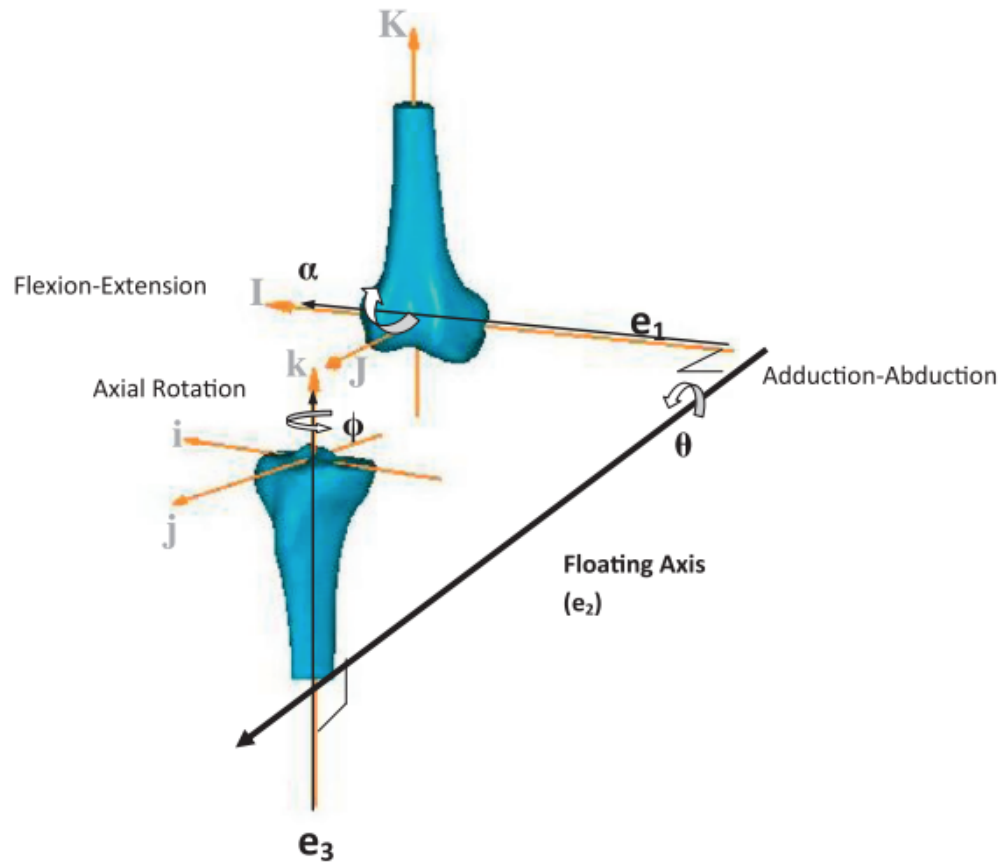
Figure 2.6: Anatomical knee rotations showing flexion/extension about the femur horizontal axis, internal/external (axial) rotation about the tibial vertical axis, and adduction/abduction about a floating axis between two segments [54].

Three anatomical rotations are defined as:

1. flexion/extension: rotation about the fixed horizontal femur axis,

2. internal/external: rotation about the fixed vertical tibia axis,

3. and abduction/adduction: rotation about a floating axis. (Figure 2.6).

In the current thesis, abduction/adduction rotation will also be noted as a change in varus/-valgus angle.

2.7   Functional Testing

Specific functional tests are effective for evaluating hip and knee OA: the stair climb, chair-stand, six-minute walk, and timed-up-and-go (TUG) tests [22]. Executing these constrained tests is beneficial for patient-patient comparison to determine normal test patterns. Traditional execution of these tests provides a start to end measurement (time taken, distance travelled, number of repetitions) but fails to identify information about joint usage intermediately. By using only a complete execution test measurement, there is no insight offered into how the specific test was completed. A patient hopping on their dominate limb or heavily compensating for an injury may record similar test measurements despite completing the tests vastly different than someone considered to be healthy and these tests would be indistinguishable from the test data alone. Additionally, parameters such as active flexion range are important indicators of knee OA whereas hyper-extension has been associated with negative outcome measures and these are not currently measured with the traditional metrics [23]. The TUG test is of additional interest because it combines the chair-stand and walking activities and also includes a turning action (or pivot). Patients begin the test seated and must rise from a chair, walk three meters, and return to their initial position. Execution time is much shorter than the six-minute walk test as it typically takes between 10-30 seconds and requires only a three meter stretch of space and a chair to implement. A recent study has revealed that TUG performance before surgical intervention can be used to predict patient length of stay (LOS) following total hip replacement (THR) or TKR and short-term functional recovery as well [24]. Upon visual observation it can be seen that patients execute this test differently by using many different strategies. To accurately measure the test, patients must be instrumented with a system capable of producing the described measurements for appropriate evaluation of knee patients.

Functional performance tests permit structured methods of assessing patients. After consulting 138 clinicians and researchers across 16 countries, three tests have been suggested by OARSI to assess the functional performance of hip and knee OA patients that can also be easily deployed in the clinic with little additional equipment [55].

2.7.1 Thirty Second Chair Stand

The chair stand test involves timing a patient while they repeatedly rise and sit from a chair. A transition from the sit to stand pose places strain on both knees and the number of repetitions can be indicative of balance and function while stressing the joints.

2.7.2 Six Minute Walk

A six minute walking test requires the patient to continuously walk for an extended period. The amount of distance travelled is the traditional test metric. A limitation of this test is that the person being studied must be able to walk continuously without being interrupted by the environment. This can sometimes be difficult in a clinic or hospital setting where there are other obstacles such as people and equipment moving about the halls.

2.7.3 Timed-up-and-go

The timed-up-and-go test combines several joint stressing activities and can be completed in a more confined location. On instruction of the observer, patients must rise from a starting seated position into a standing pose before walking three metres to a goal. Once reaching the goal, they must perform a turn or pivot before walking an additional three metres to return to a seated position in the initial starting chair.

Patients who improve on functional tests from before to after TKR are 6-8 times more likely to be satisfied with the surgery [56]. The traditional measures of these tests are number of repetitions or time to complete. While these metrics have been shown to correlate to positive outcomes following OA interventions, these remain very limited. For example, if a patient has knee OA in their left knee and has undergone a TKR, it would be expected that they may have a slower test time while healing. However, test time alone does not indicate how their operative knee is functioning, and a slow test time could be a result of an additional surgery, contralateral knee/hip OA or other mobility impairments. Using time or repetitions alone cannot distinguish a patient who may execute the tests quickly but with little mobility compared to patients who exhibit a large joint range of motion but may move more slowly. When completing tests, patients should be instructed to move at their fastest comfortable and

safe pace as to not encourage injury or falls but this also can permit some patients to push themselves more than others, despite possibly being able to perform better than demonstrated.

2.8 Machine Learning

As abundant subject data becomes available, it becomes increasingly difficult to observe patterns to make predictions or draw conclusions about future subjects. Machine learning is an application of artificial intelligence used to parse and uncover patterns in data. In general, a computer algorithm finds optimal methods of fitting or separating data. These algorithms are becoming increasingly popular in many fields including sales, marketing, and almost all other data analysis fields. Relevant advertisements can be targeted to consumers based on individualized data collected such as purchase histories and online browsing. As an abundance of digital data are available, parsing and analyzing becomes more difficult and machine learning has been shown effective when dealing with large data sets. Before applying a machine learning algorithm to help solve data problems, individual observations (or samples) must be described appropriately. Traditionally these samples are described using a number of features.

2.8.1 Features

A feature is a measurable characteristic or property of the entities under examination by machine learning models. Each sample or occurrence of data must be able to be described using the same features to use the data most effectively. Features must be numerical but there are many methods of encoding categorical data into a single value. Features can be selected manually by observing the data but in tasks such as signal or image recognition, more complex strategies can be used to automatically select useful descriptive features from data without human intervention.

In order to develop machine learning models, samples are used to create rules for classification and prediction by adjusting feature importance and developing a relation to the output class. These techniques can be broadly split into two primary categories: supervised and unsupervised.

2.8.2 Unsupervised Machine Learning

Unsupervised machine learning techniques do not require labelled data to identify patterns in data. Labels are assigned to samples as ground truth values representing the correct classification or solution when they are known, but there are instances when it is not known what a data sample represents, even if it can be described using a series of features. With unsupervised machine learning, data are interpreted as classless entities without known labels that are organically separated and grouped based on their features without defining any group membership criteria. This is useful when the significance or label of samples is not known and can reveal similarities and complex patterns, especially in samples with a large number of features where patterns are not easily interpreted manually. Following unsupervised machine learning, cluster analysis can be performed to evaluate and give meaning to separated groups for further analysis. In unsupervised learning algorithms, data are partitioned such that similar items are clustered together.

K-means Algorithm

A common example of unsupervised clustering that is used in Chapter 6 of the current thesis is the K-means algorithm. First, a fixed number of clusters are chosen (k) before data are iteratively sorted into each of the k groups. No two clusters can contain the same sample, and the algorithm repeats until no more (or a tolerance of a few) change group membership and the maximum separation between the samples has been found. Each group has a cluster centroid and features of all other samples are individually compared to find the most similar centroid before assigning membership. As clusters grow in samples, the centroids are recomputed to reflect the new group centroid. The separation effectiveness can be measured by computing the total distance of all samples to their respective clusters where the total distance should be minimized.

A silhouette plot can also be used to find an appropriate number of clusters and to examine cluster separation (Figure 2.7).

The algorithm can be outlined as follows:

1. Choose the k number of clusters and randomly select that many data points to be the

clusters.

2. Compute the distance from every point to each of the K clusters and assign all samples to the group who's centroid is closest.

3. Recompute each of the cluster centroids by computing a mean of each feature, taking into account all new members.

4. Loop to #2 until no more clusters change group membership.

Many distance metrics can be computed to establish the distance from samples to centroids. In the current work, the Euclidean distance computation has been used. It is important to note that features should be standardized before clustering to ensure features vary on the same scale, otherwise features with drastic changes in measures will more heavily influence distance computations.
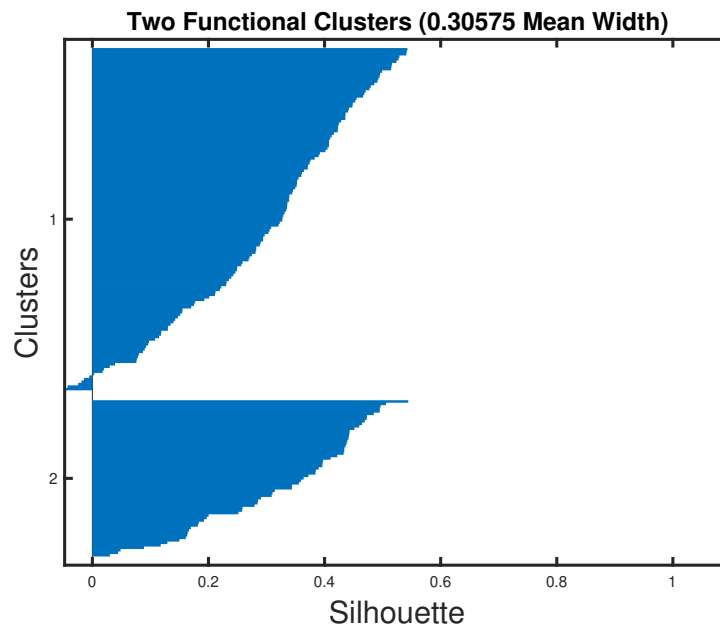
Figure 2.7: A silhouette plot showing the separation of objects into two clusters. Bars extending positively on the x-axis indicate more separation and negative bars indicate more close proximity of the samples to the incorrect clusters. A larger mean bar width indicates better separation.

2.8.3 Supervised Machine Learning

Supervised machine learning relies on a set of labelled data to develop a machine learning model. To develop and evaluate these models, data sets are split into training and testing sub-sets. Training data are used to set model parameters to find the optimal linear or non-linear combination of features that are most likely to dictate the correct output class while the training set is used at a second stage in order to test and verify the success of the trained model. An example of a supervised model used in Chapter 6 is the support vector machine (SVM). SVMs are very effective linear classifiers that work by finding the most effective separation line (hyperplane) between samples to segment sample classes. It can be seen in Figure 2.8 that there are many separation lines that can be drawn to separate classes of type "+" and class "o". An optimal hyperplane found using the training data that is separated from the class bounds will likely perform better on the testing set and future unseen data.
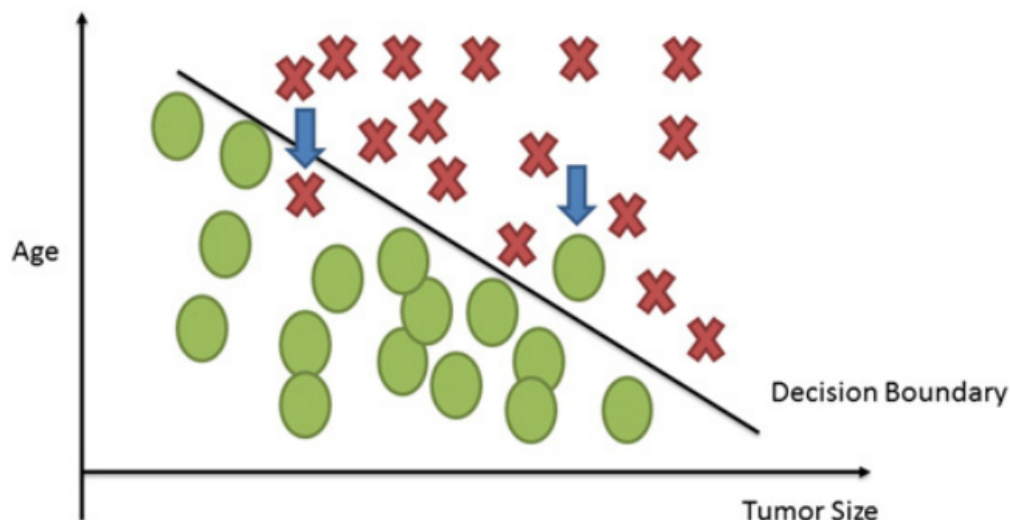
Figure 2.8: Support vectors are found to create the best hyperplane to separate classes of data [57]. Note that most machine learning problems involve many features, so the developed hyperplane will extend in many dimensions.

Although the SVM is a linear classifier, a kernel function can be applied to transform non-linear sample features into a linear space.

Naïve Bayes

This probabilistic algorithm is based on Bayes Theorem which states that the probability of A occurring can be found given B has already occurred while knowing the probability of B given A, and the probabilities of both A and B occurring Equation (2.15). To apply this equation to a classification problem, we can substitute A for the class label y, and B for the feature set $X = (x_1, x_2, ..., x_n)$.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}, \text{ or } P(y|X) = \frac{P(X|y)P(y)}{P(X)} \tag{2.15}$$

$$= P(y|x_1, x_2, ..., x_n) = \frac{P(x_1|y)P(x_2|y)...P(x_n|y)P(y)}{P(x_1)P(x_2)...P(x_n)}$$

A strong assumption of feature independence is made in Equation (2.15) when substituting in the feature set. Despite this assumption, the Naïve Bayes (NB) classifier has been shown effective in applications such as spam filtering where features such as word counts cannot be considered independent [58].

Decision Trees and Random Forests

A decision tree (DT) classifier is type of model capable of distinguishing classes that are not linearly separable. The flow of a tree begins at the top root node and branches downward as conditions are satisfied (Figure 2.9). At each step, samples are further refined until they reach an end node with the an appropriate class label.

Overfitting is a common problem with DTs, as many branches can be developed to perfectly fit all branching scenarios in the training data. This can be avoided by restricting the depth and of branches. These decisions are not usually optimized by the algorithm and must be decided before training. These fixed parameters are called hyperparameters. Often, the training data is further split into two separate training and validation sets. The validation set is used to determine the correct hyperparameters by first training model parameters and evaluating model performance using the validation set. After changing the hyperparameters, the model is retained and performance is compared using the validation set. Once the optimal set
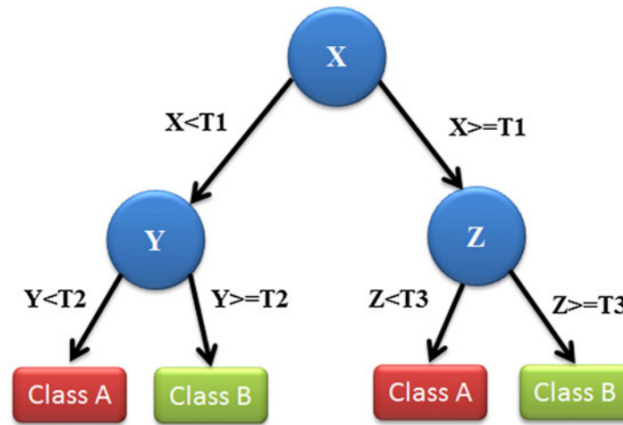
Figure 2.9: A decision tree first branching on the variable X with two future branches on variables Y and Z [57]. Parameters T1, T2, and T3 are adjusted during model training using the training set.

of hyperparameters have been chosen, the final model is trained and tested using the testing set to obtain a measure of how the model will perform on future unseen data. Hyperparameters are not unique to DTs and this strategy is used for other models with tunable hyperparameters. Validation and training sets are often shuffled as well to ensure the validation data is representative of the training set.

A random forest (RF) model is an assortment of decision trees. Often a single tree cannot properly model data without overfitting becoming too specific. An alternative method is to generate many smaller trees and take a polling approach where the most commonly classified label among all trees becomes the prediction.

Artificial Neural Networks and Deep Learning

An artificial neural network is a non-linear classification model that provides a mapping from feature inputs to outputs through hidden layers of nodes (named neurons after the human brain). In Figure 2.10, the left-most input layer matches the number of features in a sample and the right-most layer matches the number of output classes.

Arrows connecting layers indicate data travelling through from inputs to outputs that are combined at arrow intersecting nodes. Each arrow has a weight parameter that is trained to optimize the previous node's influence on the next combination. Each node also has a trainable
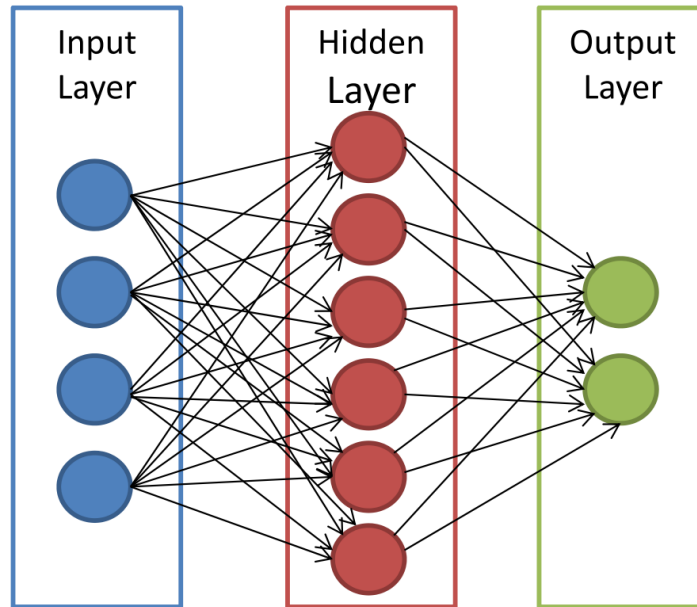
Figure 2.10: A simple artificial neural network consisting of one input layer, one hidden layer, and a final output layer. Arrows indicate that all nodes are fully connected. [57].

bias parameter. The linear combination of all inputs and their weights, and the node bias, is passed to an activation function to clamp the node's output. Several activation functions have been proposed but the Sigmoid and Rectified Linear (ReLu) functions are common choices.

Deep learning is a topic extending from ANNs. Networks are considered deep as many hidden layers are added. Additionally, each hidden layer can contain large numbers of nodes. The number of hidden layers and nodes in each layer are hyperparameters that must be manually tuned using validation sets. Training deep networks also requires significantly more data. As networks increase in the number of layers and nodes in each layer, the number of trainable parameters increases exponentially.

Convolutional Neural Networks

Convolutional neural networks (CNNs) are a specific application of deep learning focused on classification and recognition of signals or images. Face and object detection, medical diagnosis, and self-driving cars are some areas of computer vision that are rapidly advancing due to CNNs and image classification. Chapter 7 explores the use of a CNN for human activity recognition using methods similar to imaging tasks.

A CNN is a type of neural network that permits the extraction of abstracted information from images. Contrary to the models previously outlined in this section, CNN's automatically extract features of images that are useful for classification. As inputs, they take two-dimensional pixel information but these inputs can be expanded across many channels, such as in colour images where three channels of pixels can be used as inputs with each channel representing red, green, and blue pixel brightness.

A convolutional layer consists of correlating a two-dimensional filter across an image channel. Filter sizes can range depending on the resolution and level of detail of the image, but 3x3 is a common first choice. At each step of the correlation, the dot product of the filter elements and the underlying image pixels is computed. After correlating every pixel in the image channel, a new two-dimensional image is the result. Since filters are larger than a single pixel and they are correlated about the edges and corners of an image, maintaining the original image dimensions is often accomplished by padding the filter overhang with zeros. This is a design choice and non-padded, smaller images can be maintained and fed into future network layers if desired. Correlations can also be performed sequentially but non-consecutively by increasing a stride length to correlate every second, or third pixel for example. Increasing stride length will also decrease the output image size as fewer correlations are performed.

Each element of filters are tunable parameters which are optimized to best extract information from image channels that aid in successfully classifying the image. A single filter will often not provide sufficient information to classify an image and often many filters will be trained simultaneously which will also generate an increased number of output channels. Since even a small 3x3 filter contains nine trainable parameters, as the number of filters and channels increases, the number of parameters grows exponentially as well. Because of the large number of parameters, a very large number of input images are needed to train CNNs in most applications. The number of convolutional layers and number of filters varies greatly upon the intended application and detail of input images.

Techniques such as pooling can be used to reduce the dimensionality of image layers to help reduce the number of parameters. Max pooling can be included between convolutional layers to reduce a small array of pixels into a single maximum value. Average pooling operates similarly but uses the mean average of an image channel slice. The inclusion of pooling layers

and their effective width and height are additional hyperparameters that must be optimized using validation sets.

Following automatic feature extraction, resulting image layers are flattened into a single one-dimensional feature vector and are joined into a shallow artificial neural network. Often this classification network is one or two fully connected layers but the design can also vary with the application.

Chapter 3

Literature Review

This chapter contains a review of relevant works building to the work presented in this thesis. The first section explores advances in wearable technologies associated with knee measurement and proceeds into works concerning patient instrumentation and functional testing. Prediction of patient outcomes is explored next before discussing existing work on human activity recognition using wearable sensors for functional analysis in more unscripted environments.

3.1   Wearable Sensor Instrumentation

Methods of functional evaluation vary, and can include specific measurement tools to target areas of the body or more general measurements of activity. Activity monitors have become easily and readily available to the general public in many varieties. These devices include pedometers, accelerometers or full wearable systems designed to provide user feedback on exercise and activity levels.

A previously compiled report compared many common activity monitors used to monitor patients suffering with hip or knee OA [59]. The devices are shown to vary greatly in cost, hardware configuration, sample rate, interfaces, and data format. It was speculated that a sample rate as low as 25Hz can be used to capture the movement of lower extremities [59]. Most interestingly, the devices are mounted to patients in many different ways: from ankle fixtures to wrist mounted sensors. Of the 14 devices specified in the report, none benefit from mount-

ing a sensor above and below a patient's knee; the primary joint of interest. Systems are most commonly accelerometer based, using known acceleration pattern matching. Without estimating the orientation of each shank independently, certain metrics such as the amount of flexion in the knee while a patient is in an unexpected orientation such as laying on their back, for example, are not seen. Such movements may be common during physical therapy exercises and instrumentation of these sessions may be beneficial. Implementations with a single sensor do not provide sufficient information to identify knee asymmetries or changes in joint angles crucial to identifying knee-specific metrics.

In addition, IMU gait analysis solutions have been marketed to instrument functional testing of patients in a clinical setting [60]. These products fail to accurately measure knee flexion for unconstrained movements in patients with knee OA, which has been identified as a major usage characteristic contributing to patient satisfaction. It can be expected that accelerometer-only based systems using pattern matching techniques without recording independent orientations will fail to capture constant joint motion during slow activities or range of motion (ROM) exercises.

Systems specifically deriving joint angles from inertial units by combining accelerometer and gyroscope data currently exist [61–68]. The existing standard approach is to calibrate raw accelerometer and gyroscope data by performing procedures that require the patient to perform movements constrained to a single degree of freedom. During the constrained movements, the sensor frames are aligned to the attached body segment. It has also been shown that a known kinematic model of the joint can be used to measure angles between segments at high speeds by solving a non-linear observer problem. [68]. This method only requires that the sensors be at rest for alignment using a gravity vector. *El-Gohary et al.* show that this method is accurate (given several user specified parameters) through robotic testing. By contrast, the current work introduces a system which is not dependent on the attitude estimation algorithms used but relies only on quaternion sensor outputs without needing a kinematic model, reducing system computation drastically. Additionally, the world-coordinate readings readily provide information of a segment's position in the world to identify laying, standing or body turn directions during functional activity without more complex computation of body kinematics in future implementations. It should be noted that many existing systems are only tested on

healthy subjects, mostly because they are early in development, and results could vary when testing on unhealthy subjects [45]. Patients recovering from surgery may find these calibration procedures difficult. Many systems require the calibrations to be performed before every data collection session, every time the sensors are removed and replaced, or as frequently as possible [61–63, 66].

The proposed method of using independent attitude estimations from MARG sensors with respect to a global coordinate systems allows the sensors to be placed without any additional calibration before measurements can be obtained. Since each sensor is responsible for its own attitude estimation, computing the difference in orientation is efficient and can be implemented using low-cost hardware.

3.2 Performance Testing with Knee OA Patients

To obtain consistent performance measures under comparable conditions, scripted activities or performance tests can be used. By having subjects perform the same joint stressing activities across multiple instrumentation sessions, it is easy to observe areas of improvement or decline opposed to evaluating on different criteria at each observation. Furthermore, this strategy of scripted evaluation permits the comparison of different subjects since they are performing the same activities.

Specifically to the knee OA population, activities stressing the knee are preferential such as ascending/descending steps, sitting and standing, turning/pivoting, or walking/running. The degree to which these can be completed will vary drastically depending on the advancement of the OA or stage of recovery following treatment or intervention. Some patients with severe OA are unlikely to be able to climb or descend more than a few steps before becoming fatigued or experiencing pain.

The Osteoarthritis Research Society International (OARSI) have identified the utility of a core set of physical performance tests for selection of the most feasible and useful tests in hip and knee OA populations [55] Further considerations into the additional environment requirements for evaluation must also be considered. Portable steps are cumbersome and may be difficult to implement in small clinics or exam rooms while the use of full stairs may increase

the risk for injury during testing if subjects are very unstable due to mobility issues.

The timed-up-and-go (TUG) test is an easy-to-execute test with standardized guidelines; it only requires a chair, a three-meter space, and a timer. Instructions for this test are few and can usually be explained easily. Despite its simplicity, the TUG test combines many activities such as sitting, standing, turning, and walking to evaluate accumulate performance on a selection of activities. While originally developed to obtain a measure of balance in elderly subjects, the test has been shown to correlate with other important parameters of health.

The TUG test has been shown to be particularly useful for assessing the knee, especially in patients with osteoarthritis (OA) and has been included by Osteoarthritis Society International (OARSI) in a list of five recommended performance-based knee and hip functional tests [69]. The reliability of the TUG test for patients with varying severity of knee OA has been examined in literature, indicating excellent re-test reliability for both mild and moderate disease stages and a minimal detectable change (MDC) as low as $1.10s$ when timing sequential tests with the same evaluator [70,71]. It was also shown that subjects with confirmed OA performed worse on average than undiagnosed, potentially new cases ($14.3s$ versus $8.5s$) indicating results reflect progression.

Furthermore, a recent study has revealed that TUG performance before surgical intervention can be used to predict patient length of stay (LOS) following total hip replacement (THR) or total knee replacement (TKR) and short-term functional recovery [72]. The ability to predict cases of increased LOS could help plan post-operative resources appropriately or offer insight into realistic post-surgical expectations. It has been shown that the TUG test is more effective at predicting outcomes of high-risk patients than gait speed (walking) alone, indicating poor outcomes may be better predicted using the TUG [73].

Despite many useful correlations between traditional timed measurements of the TUG test and knee surgery/disease parameters, a start-to-end time measurement does not reflect how patients complete the test. Visual observation of different patients with knee OA completing the test reveals an abundance of different strategies. Timing of the entire test is not sufficient to identify effort expended, execution method, gait asymmetry, turn direction, or more specific temporal metrics derived from each individual test component. This information could be obtain by completing these functional tests in a gait analysis or motion tracking environment,

however, these systems are not commonly available and are inaccessible most patients.

Quantitative measurements are favoured over qualitative but can be difficult to obtain efficiently with these traditional measurement techniques. Recent wearable systems have been shown to provide reliable patient measurements and can be deployed in ambulatory environments without significant set-up time to instrument patients during functional testing. Since the TUG has already been identified as an important test to capture the overall function of an OA patient, applying wearable sensors to obtain more granular and objective metrics should help reveal positive or negative functional traits.

Previous work has explored the instrumentation of this test to detect functional differences in other health areas. *Mariani et al.* show that a single inertial sensor mounted on the shoe recording accelerometer and gyroscope data can be used to derive gait information during the TUG test. Gait cycles were detected as a positive peak angular velocity and were then classified as either transition or steady cycles by examining the system response time as well as turning cycles by reaching exceeding a turning angle threshold of $20°$. Stride length and velocity were computed as well from the detected cycles [74]. *Greene et al.* have also used inertial sensors attached to the left and right shin to determine angular velocity of the medio-lateral shank and used this information to increase the accuracy of fall prediction [75].

Salarian et al. propose a TUG system comprised of seven inertial measurement units (IMUs) recording angular velocity of body limb segments and the trunk [76]. A novel mathematical model for detecting turning during the test using the trunk twist as opposed to angular velocity is proposed. This is useful because slow or fast turns can be detected irrespective of local noise or sharp changes in velocity. Additionally, range of motion and acceleration of the trunk were used along with limb angular velocities to segment the test into: sit-to-stand, steady-state gait, turning, and returning to sit segments. It was found that gait was the most reliable sub-component with high intra-class correlation coefficient (ICC) for cadence and stride velocity. Extending this work, *Nguyen et al.* propose a 17 sensor system capable off segmenting the TUG test for Parkinson's patients with very high accuracy using sensor angular velocity and acceleration.

TUG tests have been used to assess an assortment of impairments and to evaluate function in elderly populations. Total test completion time has been shown to predict fall risk and can

provide an overall measure of function, however, more recent instrumented iterations of the test have proven more useful. Lightweight and unobtrusive wearable sensors have permitted many sensor configurations to be applied to instrument different body segments as applicable. Furthermore, different strategies have been explored to extract relevant data from the test. Commonly, the entirety of the recording is split into test sub-components, and from these individual sections, separate metrics are extracted. This segmentation is helped by the test's scripted structure, where the series of activities to be completed during the test is always known. Even with a rigid script, different motion thresholds can be used for the detection of the start and end of each sub-activity.

TUG tests are commonly segmented into the six sub-activities that compose the test: sit to stand, walking to the goal, turning 180 degrees about the goal, walking back to the start position, turning again, and a stand to sit. Traditionally these could be determined by a trained observer however variation between individuals makes them difficult to distinguish. For example, some subjects begin to turn to sit while they appear to be walking and combine activities into transitions. Similarly, some subjects make very wide turns while others pivot more quickly about one leg. There is no generally accepted definition for the start and end of each subactivity which makes it difficult to validate segmentation and data extraction algorithms. Results can be verified against human observers however these are often far from ground truth labels. Existing literature from other applications of instrumented TUG tests provides a base to expand on.

A common sensor mounting location is the torso where the sensor can measure the body's centre of mass. *Higashi et al.* instrumented ten subjects with a torso and thigh sensor configuration and segmented their TUG tests into six sub-components. The standing activity was detected when the thigh sensor pitch (axis outward from hip laterally) exceeded a change of 10 *deg* from the seated pose, walking detected when the pitch exceeds 10 *deg/s*, turning detected with a large angular velocity signal from the waist sensor (determined experimentally), an additional walk period before turning once more, and sitting when the thigh pitch becomes less than 10 *deg* from the test start pitch [77]. Segment times as well as acceleration of the trunk and thigh were used to further analyze subject motion.

Nguyen et al. confirmed that automatic detection of individual segments of an instrumented TUG test was more reliable than independent observers by instrumenting 16 healthy subjects

while completing a modified 5m and 10m TUG test [78]. Sensing equipment consisted of a full-body suit containing 17 inertial units. Tests were divided into six segments of activity: stand up, walk-out, turn, walk-in, turn, and sit down using motion thresholds of applicable sensor locations. Transitions were also identified by detecting points before and after each activity. Raw accelerations from sensors were used in addition to orientation measurements output from a sensor fusion algorithm. These data were retrieved as quaternions which were used to compute relative motion between two sequential timepoints from which absolute range of motion metrics were extracted. This segmentation accuracy was confirmed by two observers by comparing an animated avatar reconstructed from subject motion to test video footage.

While a full body-worn configuration has been shown very effective for analyzing the TUG, it is not practical for standard of care use in the clinic for instrumenting many patients in a limited time. More minimal solutions have been explored including work by *Ortega-Bastidas et al.* which uses only a single torso-mounted inertial sensor. In this work, a stand activity is detected by a change in pitch between the subject's torso and the backrest of the starting chair. Turn starts and ends are detected by observing the maximum and minimum rate of change of normalized yaw angles of the torso (about the longitudinal anatomical body axis) [79]. Results of test segmentation are verified against visual inspection recorded with a camera-based motion capture system and shows strong agreement. Errors are small when detecting a sit-to-stand to walking transition near the beginning of the test (mean 0.07s) but errors are larger when detecting the walking to stand-to-sit transition near the end of the test (mean 0.29s) [79]. Authors of the current work suggest that this could be in part due to the varying turn-to-sit strategies observed when subjects complete the TUG which makes it difficult to determine an exact transition point with visual observation.

Subject turning identification presents a challenge in both scripted and unscripted instrumentation sessions. Especially in subjects with movement impairments, a turn may be comprised of several small and segmented turns or single turns at varying speeds. Turning is a major element of mobility and difficulties are a major contributor to falls and reduced quality of life in elderly populations and for people with movement disorders [80]. With unscripted monitoring, algorithms must differentiate between a series of small turns and continuous large turns without additional information concerning the attempted task. This is simplified greatly

through instrumentation of the TUG since the activities attempted by subjects are already established. *El-Gohary et al.* have used a similar torso-mounted inertial unit to instrument healthy subjects and those with Parkinson's disease while analyzing their turns. Using rotation of the torso (measured by angular velocity of the sensor) they were able to segment turns with a high degree of accuracy but it was discussed that some disagreement between sensor segments and observer ground truths could be due to conflicting definitions of turns. Observers watching video of the turns were instructed to note a turn as the heel strike prior to a turn and at the end while it is noted that subjects often turn their pelvis ahead of their feet when turning naturally, highlighting the unreliability of ground truth comparisons [80]. It was also noted that movement impaired subjects had slower turning velocities as would be expected for the study population and also that these patients appear to walk better when examined in the clinic opposed to caregivers' reports about daily activity, demonstrating an observer effect. The authors were able to derive metrics that could differentiate between healthy subjects and those with impaired movements. Compared to the control group, patients with Parkinson's disease took shorter turns with sharper angles but required more steps to accomplish turns.

3.3 Predicting Patient Outcomes

A variety of outcomes following total knee replacement (TKR) are possible. Entering the surgery, patients may be drastically different ages, have varying BMI, functional abilities, and outcome expectations. The ability to observe pre-surgery information and predict possible outcomes such as length of hospital stay, number of physiotherapy sessions to be required, chance of infection, or satisfaction would be very useful to appropriately allocate resources needed to ensure the best surgical outcomes possible on an individual level. At the very least, providing surgeons with more information concerning how the patient's joint performs may help influence surgical decisions and implant choices. It has been thought that parameters such as age and BMI may indicate surgical success chances, however, recent literature shows that functional performance may be a truer indicator. Research has demonstrated that traditional timed-up-and-go (TUG) test results can be used to predict length of hospital stay post-surgery [72]. With significantly more information available with the TUG test extensions described in

the previous chapter, there may be more opportunities to predict outcomes following surgery based on the functional performance measured pre-surgery or in the follow-up appointments shortly after. Identifying parameters or milestones important to successful recovery will help clinicians identify and treat problematic recoveries. Preoperative range of motion (ROM) measured passively is known to predict postoperative ROM. It is expected that ROM during weight-bearing activities may also indicate postoperative ROM so extraction of this information will be further explored in Chapter 5.

Previous work has investigated predicting pain one year after TKR by evaluating patients pre-surgery and at important time points following surgery using patient questionnaires and functional tests [81]. It was shown that preoperative pain was an indicator of postoperative pain and suggested that pain reduction in the weeks following surgery may indicate pain will continue to be reduced continuing through recovery. Early detection of patients without this improvement could be beneficial for recommending physiotherapy programs designed to help reduce future pain.

Total knee revision is a costly surgery involving the removal and re-insertion of an implant. Certain PROMs have been identified to help predict the likelihood of revision following some types of TKR. It was also discussed that although preoperative PROMs can be used to predict postoperative measures, patients with severe impairment preoperatively seem no less satisfied following surgery [82]. Additionally, patients with less severe disease progression before surgery are at a higher risk for dissatisfaction following the joint replacement since their functional improvement is less substantial. Generally, when a surgery is performed on patients with progressed stages of knee OA and severe impairment has occurred, it is expected that functional improvement expectations following surgery should be lower [83]. If more quantitative functional data is acquired from patients pre- and post-surgery, there may be a chance to identify important parameters that can predict patient outcomes, or indicate the best opportunities for intervention to improve these outcomes along the recovery path.

Machine learning for predicting outcomes has been used in other areas of the health care industry, particularly where disease outcome is the most difficult to predict on an individual case by surgeons. While algorithms can be used to help health professionals make more informed decisions, ultimately individuals are unique and must require diagnosis validations from clin-

icians. For example, a variety of personal and derived features have been used with support vector machines (SVMs), artificial neural networks (ANNs), and decision trees to aid cancer prognosis [57]. A comparison of learning algorithms to predict severe complications (sepsis and acute kidney injury) following major surgery has been compiled [84]. The data set of pre-/postoperative data had 285 features extracted and outcomes were determined using standard medical evaluations. Four learning models were tested: Naïve Bayes, generalized additive model (GAM), logistic regression, and SVM with the logistic regression and GAM performing best on the data used. Machine learning has also been applied to the prediction of subjective patient outcomes. Preoperative outcome subjective measures combined with demographics and medical history were used to predict post-surgical satisfaction using several models but it was found a Random Forest (RF) model performed the best [85]. One-year post-surgical outcomes of sensor-measured gait parameters have also been predicted using information from a single foot-worn sensor [86]. It was found that high functioning patients were more likely to see functional performance decreasing after one year of recovery.

3.4   Human Activity Recognition

Wearable inertial sensors have permitted extensive human data collection beneficial for both personal health monitoring and research tools. Accelerometers and gyroscopes are being manufactured with lower cost and smaller physical size which permits their inclusion in many consumer products. The availability of this hardware combined with the prevalence of mobile devices for data collection have fuelled a new industry for personalized health and activity monitors allowing individuals to track their fitness levels, measure functional performance, and set goals for better health. For unconstrained subject evaluation, wearable systems have been favourable over camera-based alternatives which either require tethering to an environment, or involve extensive home monitoring setup.

Patients have reported to prefer lightweight and non-detectable systems attaching with a band for easy attachment and removal instead of more permanent solutions [87]. Inertial-based wearables commonly consisting of one or more accelerometer(s), gyroscope(s), or magnetometer(s) have previously been investigated and determined appropriate for patient instrumentation

[88].

Minimal systems for long-term, energy efficient activity surveillance have been used to monitor subjects remotely using wrist, torso, thigh or chest mounted devices [89–92]. Data recorded using only accelerometers can be analyzed to identify periods of rest/activity and in some cases can be used to distinguish the activities performed by the instrumented subjects [89]. Tri-axial accelerometers can be used to obtain better estimation of subject postures or instrumented body segment orientation than their uni-axial alternatives. Accelerometry devices can also be used to accurately detect intensity of activities by evaluating the magnitudes of acceleration signals which may be useful for assessing overall levels of health, but does not give any performance information for specific activities unless further classifications are performed [91].

Alternatively, mobile phones containing accelerometers have been used for activity classification to distinguish activities including walking, running, ascending/descending stairs, sitting and standing [93–95]. Kwapisz *et al.* windowed samples into 10s intervals and six signal features were manually extracted prior to classification. Logistic regression (LR), artificial neural network (ANN) and decision tree (DT) classifiers were compared. Validation was performed using a ten-fold (n-fold) scheme where all recorded samples are combined and a testing set is withheld for algorithm evaluation. The ANN outperformed both the LR and DT classifiers. Since the number of samples varied greatly across each class, a weighted average of classification accuracy was used with an overall measure of around 92%. The accuracy when classifying ascending (61.5%) and descending (44.3%) stairs was substantially lower than the average across all classes. Voicu *et al.* used similar sized windows of smartphone measured motion and manually extracted five features for classification. An ANN was deployed with similar accuracies. The most misclassified samples were ascending/descending stairs and it was discussed that because the window was 10s, these samples also included flat-level walking where participants bridged two flights of steps which may account for some of the confusion.

Wrist-worn wearables provide little information concerning the lower extremities, however, arm activity can estimate walking distance or distinguish physiotherapy exercises [96,97]. They have been explored for classifying daily living activities of lying, standing, sitting, walking, running, washing windows, cleaning, stacking shelves and sweeping into four potential

classes: sedentary, household, walking and running [89]. Windows of activity samples were set to 12.8s and eight time and frequency domain features were manually extracted. Support vector machine (SVM), ANN and DT classifiers were compared. Performance was evaluated with multiple iterations of a leave-one-subject-out (LOSO) scheme where a classifier is trained on samples from $n-1$ subjects and the remaining subject is used for testing. A best overall accuracy of 97% was achieved using a DT, however, many activities were classified into only few categories and the chosen classes could encompass a large variety of activities with little evidence that distinction could be made between the sub-activities (ex. cleaning vs. stacking shelves). It is expected that shorter windows of activity would cause accuracy to drop as well.

Another study using a commercial torso mounted accelerometer (MoveMonitor - McRoberts, NL) has differentiate similar activities of daily living including navigating stairs and biking into classes of: shuffling, locomotion, standing, sitting or lying [98]. Walking and ascending/descending stairs were perfectly classified as locomotion but with no separation of the individual activities performed. Lying and sitting were 93-96% correctly classified but standing was misclassified as sitting 82% of the time. Cycling activities were almost entirely classified as sitting or shuffling depending on the cycle speed and not as locomotion [98].

It remains difficult to compare human activity recognition (HAR) accuracy throughout literature since implementations vary across: subject health or functional impairment, number of sensors and their placement locations on the body, activities performed, number and type of classes to distinguish, and technique validation. Additionally, various sensors may record with different measurement accuracies and this information is not often reported. When validating results of trained models or classification strategies, some papers use an *n-fold* process where samples from all subjects are blended and a portion is withheld from training to serve as a test set [98]. An alternate scheme involves leaving one subject out (LOSO) so that the test set provides an estimate of the classification strategy effectiveness when used on an unseen subject [89]. It is expected that the LOSO process is a more conservative evaluation technique and should be used for algorithm assessment since the end deployment will assess unseen subjects during future instrumentation sessions.

A more recent review by *Fawaz et al.* discusses the state-of-the-art of the more generalized problem of multivariate time series classification (TSC). It is suggested that any classification

problem using naturally ordered samples (accumulated over time, for example) can be framed as a TSC problem [99]. In addition to human activity recognition, temporal ordering of samples is present in human cognitive processes, electronic health record analysis, acoustic scene classification, and cybersecurity or intruder detection. A traditionally effective approach of comparing time series was a nearest neighbour classifier using the Dynamic Time Warping distance function. It has been shown that ensembles of classifiers can increase accuracy at a cost of computational resources [100].

Deep learning is quickly becoming effective in numerous machine learning applications and can offer an end-to-end machine learning solution free of bias induced from manual feature extraction [101]. Automatic extraction of class-discriminating features can be preferable over manual features because it is not known if an optimal set of features has been selected to differentiate specific classes. Convolutional neural networks (CNNs) and ANNs are widely adopted for TSC problems [99]. One limitation of ANNs is they exhibit spatial invariance since all sample elements (and their trained relations) are treated as a non-linear combination of independent features, regardless of where they are ordered spatially in a sample's set of features. Motivated by successes in image analysis and natural language processing CNNs offer automatic feature extraction while maintaining spatial relevance through the order of the feature set.

When applied to a univariate time series, a convolution can be seen as correlating a filter across a discrete signal. This correlation produces a transformation of the series across only the time dimension [99]. Applying several filters to a series will produce a multivariate time series intended to distinguish different features for classification success. Networks are trained using backpropagation such that filter values are adjusted so correlating filters create responses that distinguish classes of samples. Two dimensional CNNs used for image recognition employ 2D filters which have the capability of producing responses to features in both height and width dimensions of an image. This is especially useful for detecting localized features in an image, or pattern of pixel values that represent distinguishing elements. Previous literature has demonstrated a 2D CNN using raw accelerometer inputs can accurately classify hand gestures [102]. The CNN outperformed SVM, k-nearest neighbour (KNN), and deep belief network (DBN) classifiers on several benchmark datasets and demonstrates the utility of detecting responses

across the time and spatial domains of raw sensor data.

Methods of encoding time series data into images have been proposed. Wang *et al.* presents a Gramian Angular Field (GAF) method where univariate time series data are normalized on a scale of [-1,1] and transformed to polar coordinates instead of typical Cartesian coordinates [103]. Yang *et al.* extend this work for multivariate applications by aggregating each GAF image generated for each univariate signal [12]. Each grayscale GAF image is treated like an independent input data channel, similar to how a coloured image occupies multiple input channels in traditional usages of CNNs for image classification.

Given the expectation that the average human observer could classify many daily activities based only on observation of a subject's legs, the current work will employ a more direct multivariate time series encoding which maintains (and does not abstract) the spatial relation given by the raw sensor orientation readings, similar to the work of Bo Yang *et al.*. If sensor readings were to be split between separate input channels, developed filters would not respond to patterns across separate input time series.

3.5 Summary

This chapter has outlined several topics pertaining to the work that will be presented in future chapters. An overview of wearable sensor systems appropriate for subject instrumentation has been presented first. While other sensor systems exist, they are often too complex or constraining for efficient use in the clinic. Systems consisting of many sensors can also become expensive and would not be practical for mass patient deployment. As previously mentioned, complex motion capture (such as marker or camera-based systems) can also require specific environmental setups that do not translate well to existing clinic hallways or waiting rooms. The system development in the following chapter has focused on developing a knee measurement system that could be efficiently integrated into clinical workflows to quickly provide useful knee information.

The next topic concerned functionally measuring patients. Traditional outcome assessment relies on patient feedback through validated questionnaires but it has been highlighted that these subjective measures make it difficult to distinguish more subtle changes in health. They

are also influenced by patient bias making it difficult to make direct patient-to-patient comparisons. This chapter has outlined several appropriate functional tests that could benefit from instrumentation to obtain objective feedback. The timed-up-and-go test has been highlighted as it combines many knee stressing activities and can be completed by subjects relatively quickly.

Following functional testing, this chapter explored the possibility of predicting future patient outcomes after undergoing total knee replacement. Existing literature has mostly considered subjective measure predictions although sensor-instrumented measures such as gait parameters. By incorporating a well-known functional test such as the timed-up-and-go into instrumentation sessions, the current work will explore the prediction of functional improvement likelihood using traditional test measures as a proxy for functional capability.

Lastly, this chapter discussed some challenges with sensor instrumentation outside of the clinic and offered human activity recognition solutions. Extensive literature has explored the possibility of classifying well-defined sets of activities using inertial sensors but was lacking indications that these classifications could be performed using only an inertial sensor mounted above and below each knee. The current work will attempt to answer if daily activities could be classified by only observing a subjects legs. A shortfall of more traditional machine learning for human activity recognition requires manual extraction of parameters to input to to machine learning models. It was discussed that automatic feature extraction can offer better performance so the current work describes a novel strategy to encode raw sensor data into images to leverage previously successful convolutional neural networks for automatic feature extraction.

Chapter 4

Wearable Sensor System Development

4.1 Introduction

The current chapter describes the development and validation of a wearable sensor system appropriate for in-clinic instrumentation of patients following total knee replacement (TKR). A strategy for sensor placement is presented and validated using a robotic manipulator and leg phantom placed inside a 3D motion capture environment. Joint angles are compared to the gold standard measurement device through several trials while removing and replacing sensors on the robot phantom.

4.2 Wearable Sensor Considerations

Investigation began into possible wearable system types that would function best for measuring joint angles in patients with knee OA. Wearable systems consisting of tight fitting apparatus were discouraged since it was thought that irritation could occur if mounted close to an incision. Additionally, if the patient is experiencing any pain or inflammation of the joint, it would not be ideal to attach anything conforming to the area. Stretchable fabric, or smart materials, were eliminated as a useful option for these reasons. Similarly, it was a concern that using digital goniometers could cause the same discomfort attached in proximity to the incision. Additional adhesives would be required to attach the goniometers and they must remain affixed to report results of any significant accuracy. A major limitation of these devices is that they

cannot be separated and attached proximal and distal to the joint effectively because they are limited in physical stretch of the measurement wire. Despite the devices being manufactured in many possible lengths, as the length of wire increases, as does its tendency to be influenced by external forces. For example, during activity when the knee is flexing and extending, the internal measurement wire will become tensed and lax, and may be influenced by the motion of the body while performing activities. Clothing and skin motion may also cause artificial readings of joint motion in practical use.

Inertial sensors were then examined as a possibility for instrumenting patients since these devices are becoming small in physical size and more energy efficient. Wireless sensors could be attached at variable lengths from the affected joint without causing discomfort. These devices have been previously used to track joint angles. It has also been noted that non-negligible error can be introduced using previous methods of measuring joint angles with inertial sensors if no frame alignment procedure is done. These procedures require movements to be performed constricted to known rotations in order to align the local sensor coordinate frames with the attached joint segment frame. After observation of patients before and shortly after knee replacement, it was determined that the removal of these calibration procedures would be ideal. If the correct execution of calibration procedures would be dependent on the sensor system accuracy, there could be a mixed success rate using this method as patients had vastly different functional abilities. In literature proposing these methods, the procedures were all completed by healthy individuals and it was apparent when observing knee OA or TKR patients that some individuals would have a great deal of difficulty with these.

A strategy of tracking joint angles was explored by independently estimating an orientation with respect to an absolute reference by each sensor. The motivation for this method would be to allow the sensors to calibrate to an absolute reference and then when attached to a body segment, the body segment would have the same frame as the sensor. With a sensor attached above and below each knee, a difference in limb segment orientation could be used to derive joint angles in three dimensions. In addition to being able to derive these measurements instantaneously without any patient calibrations, this method would also give absolute world body orientations without the need for a kinematic model. Additionally, since each sensor implements an algorithm taking raw component measurements and computing an attitude esti-

mation, the method of extracting joint angles and analyzing the patients would be completely decoupled from the algorithms implemented on the sensors. This would allow sensors from different manufacturers to be used and as attitude estimation algorithms improve, the high level system can remain unchanged except for interfacing with the new sensors. Many attitude estimation sensors output their absolute orientations in quaternions as well, which makes computing changes in orientations efficient using linear combinations.

4.2.1 Implementation

As an input, the proposed strategy requires sensors capable of transferring or streaming quaternion attitude estimations wirelessly. MARG development boards (MetaMotionR, MBientLab, San Francisco, CA, USA) capable of estimating orientations using a self-contained sensor fusion algorithm at a rate of 25 Hz were obtained. It has been previously suggested that this recording frequency should be sufficient to monitor lower extremities outside of high-speed activities [104]. Orientations are measured relative to a global left-handed frame defined identically for all manufactured sensors with the positive *z* axis downward, the positive *x* axis towards magnetic north and the negative *y* axis orthogonal to the two former axes, calibrated in San Francisco, CA, USA. The boards and a rechargeable lithium-polymer battery were placed inside a custom PLA 3D-printed case (visible in Figures 4.1 and 4.2) approximately 1.2 *cm* thick, 3.0 *cm* wide, and 4.0 *cm* long.

An iPod application written in Swift 3 was created to facilitate the initialization and synchronization of each of the four sensors during recording sessions. The application permits users to first select an existing subject ID or generate a new one. Subject IDs are generated as universally unique identifiers (UUIDs). Since these identifiers are guaranteed to be unique, they are used to identify sessions but also as subject identifiers in future system use.

Once a subject or session has been chosen, the application first instructs a user to attach the sensors using stretchable fabric hook and latch straps to their limbs as indicated in Figure 4.1a. The system then connects to four sensors (two on each limb of a patient) and once connected, indicates that they press the momentary push-button on each sensor in a specific order to localize the sensors to a body segment. Once pressed, the system will internally identify the sensor

(a)   (b)
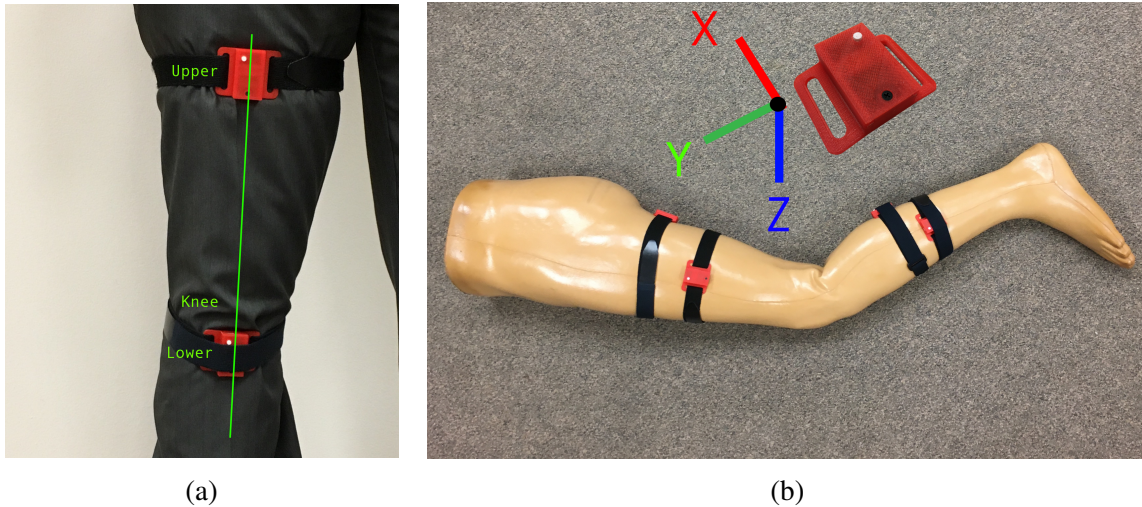
Figure 4.1: (a) Sensor attachment instruction for patients and clinicians; (b) Instrumented phantom leg (Sawbones Fully Encased Leg, Pacific Research Laboratories, Vashon, WA) with posterior and lateral placements and local sensor frame.

(a)   (b)   (c)
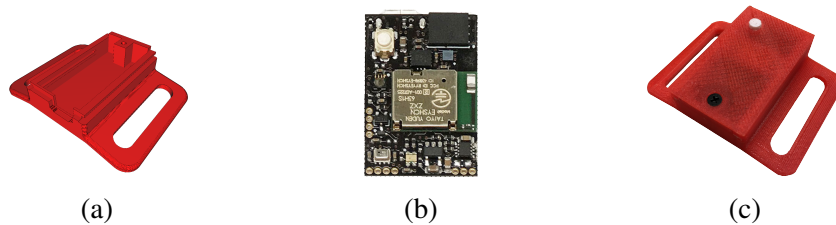
Figure 4.2: (a) CAD model of open case; (b) MetamotionR development board; (c) Final printed case assembly.

location for output and angle calculations. This interaction differs from a sensor calibration procedure because no constrained patient joint movement is required.

As joint motion is computed using the algorithm below, data are simultaneously logged to the connected portable device and synchronized. The system recognizes a disconnect or communication failure with any of the attached sensors. In order to prevent data loss in this event, quaternions are simultaneously stored to the on-board sensor memory to be retrieved by the application upon re-connection if a data synchronization error occurs from a lost connection. All data stored on-board or transferred to the portable device wirelessly are timestamped to allow merging of these two data storage methods to create a single synchronized data set.

Motion Extraction Algorithm

Once the sensors are connected to the system wirelessly and their locations have been labelled internally, the user may begin recording. The joint angles of each leg are calculated identically, but independently.

Recall from Chapter 2 that the quaternion orientation estimation read by the lower sensor with respect to its global coordinate system. It can be defined as:

$$
{}_{L}^{G}\boldsymbol{q} = [q_{l0}\ q_{l1}\ q_{l2}\ q_{l3}]
$$

and similarly we can define the orientation of the upper sensor with respect to the same global coordinate system as:

$$
{}_{U}^{G}\boldsymbol{q} = [q_{u0}\ q_{u1}\ q_{u2}\ q_{u3}]
$$

The difference in orientation of the upper sensor with respect to the local coordinate frame of the lower sensor (chosen as a reference) can be found using Equation (2.14) as:

$$
{}_{U}^{L}\boldsymbol{q} = {}_{L}^{G}\boldsymbol{q}^{*} \otimes {}_{U}^{G}\boldsymbol{q} \tag{4.1}
$$

using the lower conjugate as defined in Equation (2.10) and the linear combination defined in Equation (2.11).

The difference in orientation ${}_{U}^{L}\boldsymbol{q}$ is now the single combined rotation from one segment's local frame to the other, with respect to the reference. To maximize this experiment's clinical relevancy and present results comparably to knee literature, ${}_{U}^{L}\boldsymbol{q}$ was transformed into flexion/extension (rotation in sagittal plane), internal/external (rotation in vertical plane), and varus/valgus (rotation in coronal plane) clinical angles using a Z-Y-X Euler sequence for lateral sensor measurements and a Y-Z-X Euler sequence for posterior sensor measurements. These sequences were motivated by performing the flexion rotation first because the majority of motion occurs about this sensor axis, varus/valgus second because the knee physically cannot reach ±90 deg varus/valgus (rendering the rotation singular), and then the third remaining internal/external rotation about this transformed frame as has been done previously in literature [105]. Different sequences are required for each sensor placement despite calculating ${}_{U}^{L}\boldsymbol{q}$ iden-
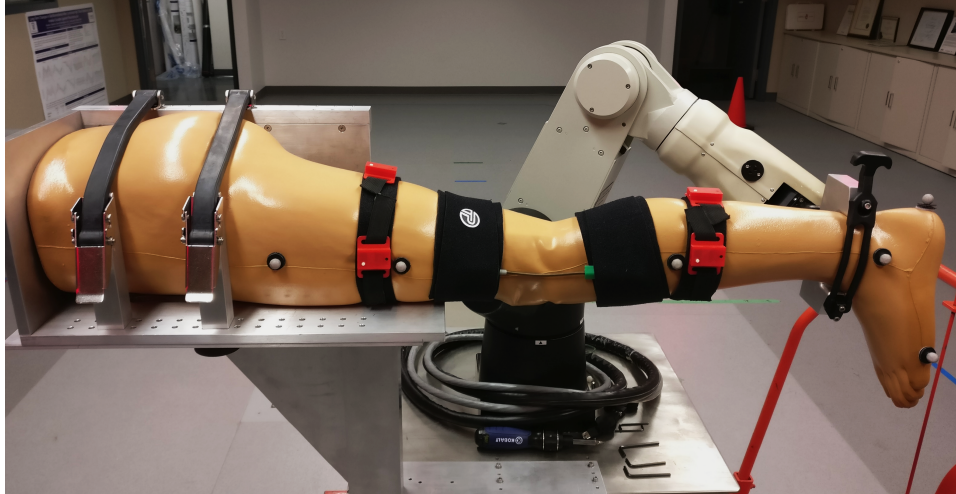
Figure 4.3: The robotic manipulator testing apparatus complete with wearable sensors and 3D motion capture markers.

tically since each sensor has a fixed local frame (visible in Figure 4.1b) and when sensors are mounted differently about the knee, motions are reflected about different sensor axes.

4.3   Methods

To validate readings obtained by the proposed system, a robot phantom study was performed. A robotic manipulator was fitted with a mannequin leg (Figure 4.1b) and was programmed to flex the leg starting completely extended (0 deg) through a linear path to approximately 120 deg for ten cycles with near-constant velocity. The manipulator apparatus holding the phantom limb can be seen in Section 4.2.1. The upper limb was secured to a stationary apparatus and the manipulator end effector was attached to the ankle. The manipulator and limb were placed inside a marker-based 3D motion gait lab (Cortex 2, Motion Analysis Corporation, Santa Rosa, USA) and fitted with reflective markers to provide a ground truth measurement for angles recorded during testing. Three independent tests were performed at two different speeds; slow ($\approx 15 \deg / s$) and fast ($\approx 25 \deg / s$). These testing speeds were limited by the maximum manipulator velocity at the end-effector. Two sensor pairs were attached in lateral and posterior positions on phantom limb as indicated in Figure 4.1b. The two simultaneous placements simulate two convenient placement options for sensors in patients with knee OA. Testing an anterior placement was not possible due to the anterior upper limb being blocked by the fixture

apparatus. Between each of the three tests, all MARG sensors were completely removed from the phantom, shuffled randomly, re-positioned on the leg, and re-localized using the application specific push-button procedure. The placement and re-placement of the MARG units between tests simulates the removal and reattachment of sensors from patients. No calibration of the system was performed between tests or trials.

4.3.1   Data Collection

Data were collected using the custom developed mobile application. For this experiment, the synchronized data files were downloaded from the iPod and post-processed using Matlab (MathWorks, Natick, MA, USA). The analysis procedure from the proposed system was implemented using a Matlab script in addition to the original mobile Swift 3.0 implementation.

4.3.2   Analysis

Clinical angles were extracted from the sensor data using the proposed algorithm. Angles for the gold standard were found as an X-Z-Y Euler sequence that labels rotations in the same order with flexion as the first stage, then varus/valgus, and internal/external rotation as the final stage. Initial offsets were subtracted from data sets to align all tests to a common start for test-to-test comparison. No further filtering was performed on the data.

A cross correlation was used to determine if the posterior or lateral sensor placements provided the most repeatable path across multiple tests with device replacement. Tests were interpolated using the Matlab $interp1$ function to normalize timestamped test data from the same axis to the longest test set. This is necessary because sensor data streaming rates varied slightly when transferring across the wireless connection and data could not be exactly compared sample-to-sample. Correlation matrices comparing all three tests were computed using the Matlab function $f() = corrcoef(A, B)$ and the correlation coefficient was taken as the matrix diagonal $d(f(A, B))$. A mean measure of correlation c for each sensor placement option and the gold standard, for each degree of freedom was found using:

$$c_{axis} = \frac{d(f(t_1, t_2)) + d(f(t_2, t_3)) + d(f(t_1, t_3))}{3} \tag{4.2}$$

where t_n is the nth test at the same trial speed.

Since test correlations only compare the retest reliability and similarity of each placement, an average amplitude for all cycles of each test was computed as $amp_{test} = mean(peaks)$ to compare ranges in each axis of both placements to the gold standard, where *peaks* is a set of all cycle amplitudes from complete extension to the cycle peak. A mean average of all tests at a single trial speed for each axis degree of freedom is found as:

$$amp_{axis} = \frac{\sum_{test=1}^{3} amp_{test}}{3} \tag{4.3}$$

To provide additional comparison of sensor placements to the gold standard, a value of root mean square error (RMSE) of each sensor test compared to its respective motion capture test was computed using the following equation:

$$RMSE(x, y) = \sqrt{\frac{1}{N} \sum_{n=1}^{N} (x(n) - y(n))^2} \tag{4.4}$$

where x, y are sensor data and gold standard data sets respectively, and N is the number of samples recorded. Note that all sensor test data sets were interpolated to gold standard tests for direct comparison using the Matlab *interp*1 function as done previously when computing cross correlations.

Finally, a measure of drift d_{test} was found by computing the change in amplitude between each successive cycle in a single test. The following Matlab equation was used: $d_{test} = mean(diff(peaks))$, where *peaks* is a set of all cycle amplitudes from start to the cycle peak. A single mean value of drift for each placement option and the gold standard was found for each axis of rotation as the mean of all three test values.

4.4 Results

4.4.1 Path Correlation

Path correlations for each placement option and the gold standard through all tests can be seen in Figures 4.4 to 4.6. The mean values calculated with Equation (4.2) are shown in
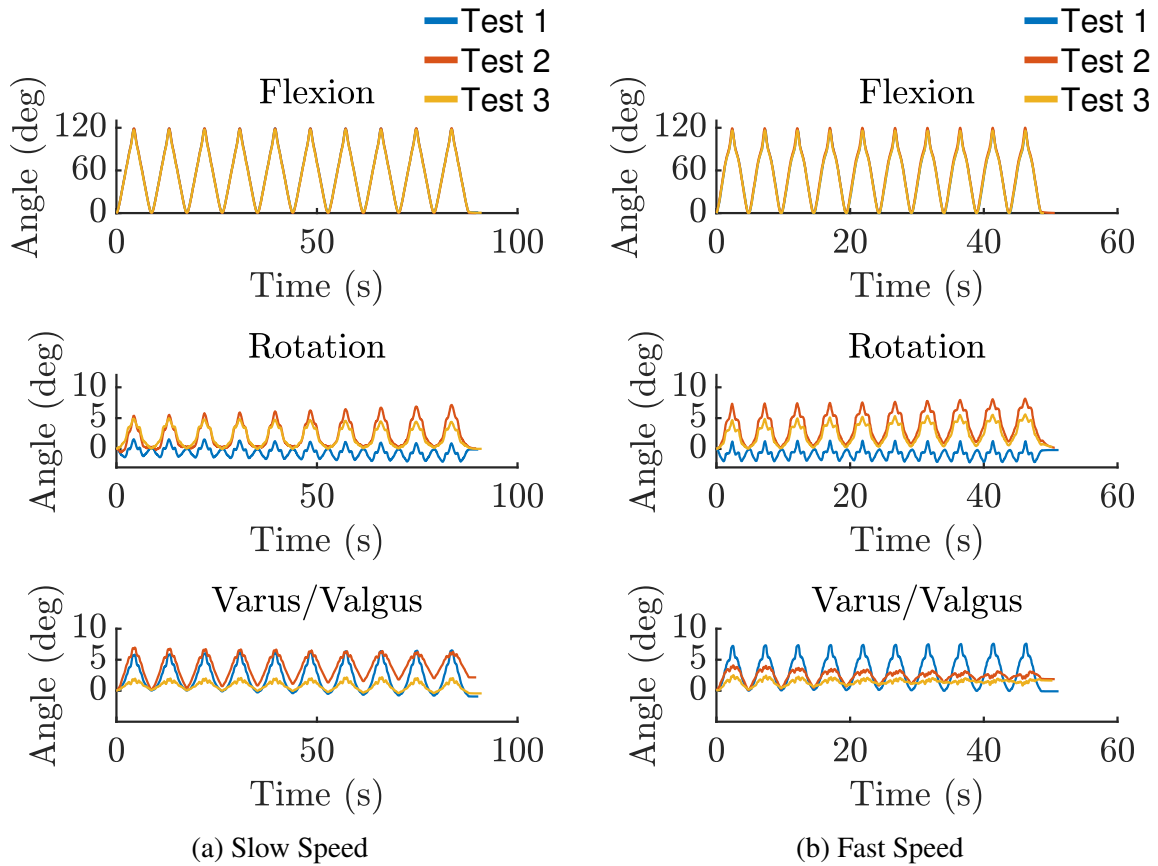
Figure 4.4: Angles computed during motion paths as the phantom limb travelled through 10 cycles using a posterior sensor placement.

Table 4.1 along with their standard deviations. It can be seen that most paths were highly correlated, with less strong results showing a large standard deviation indicating that a single test may be lowering the mean. Flexion angle path correlations were the highest (1.00 ± 0.00 rounded to two decimal places) through both placements and speeds. Rotation angles have shown the lowest correlation, however, it should be noted that the overall amplitude of these rotations is low compared to other axes (mean of all placements $\approx 3.5\ deg$) as seen in Table 4.2, so any small differences or noise in the signal have a higher impact on correlation. Similarly, flexion amplitudes are large ($\approx 120.0\ deg$) so noise and errors have a less significant effect on these path correlations.
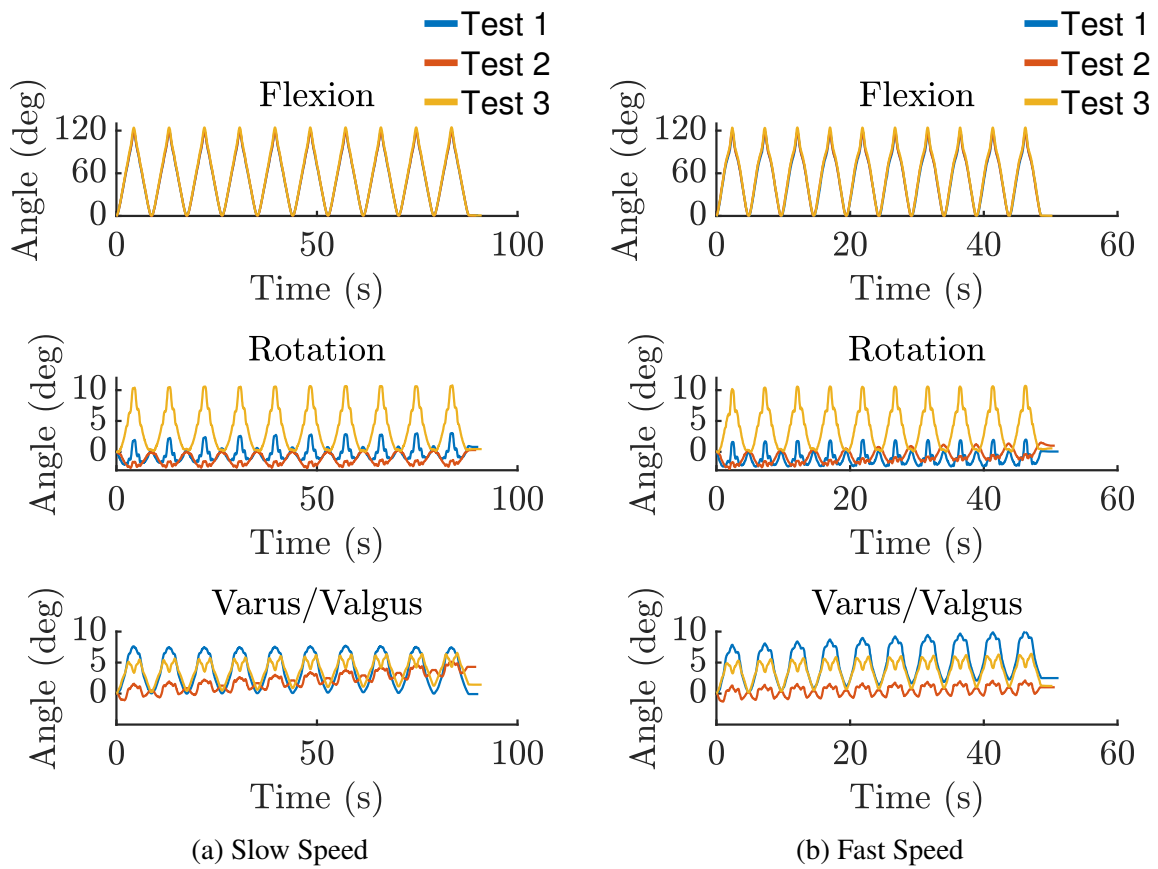
Figure 4.5: Angles computed during motion paths as the phantom limb travelled through 10 cycles using a lateral sensor placement.
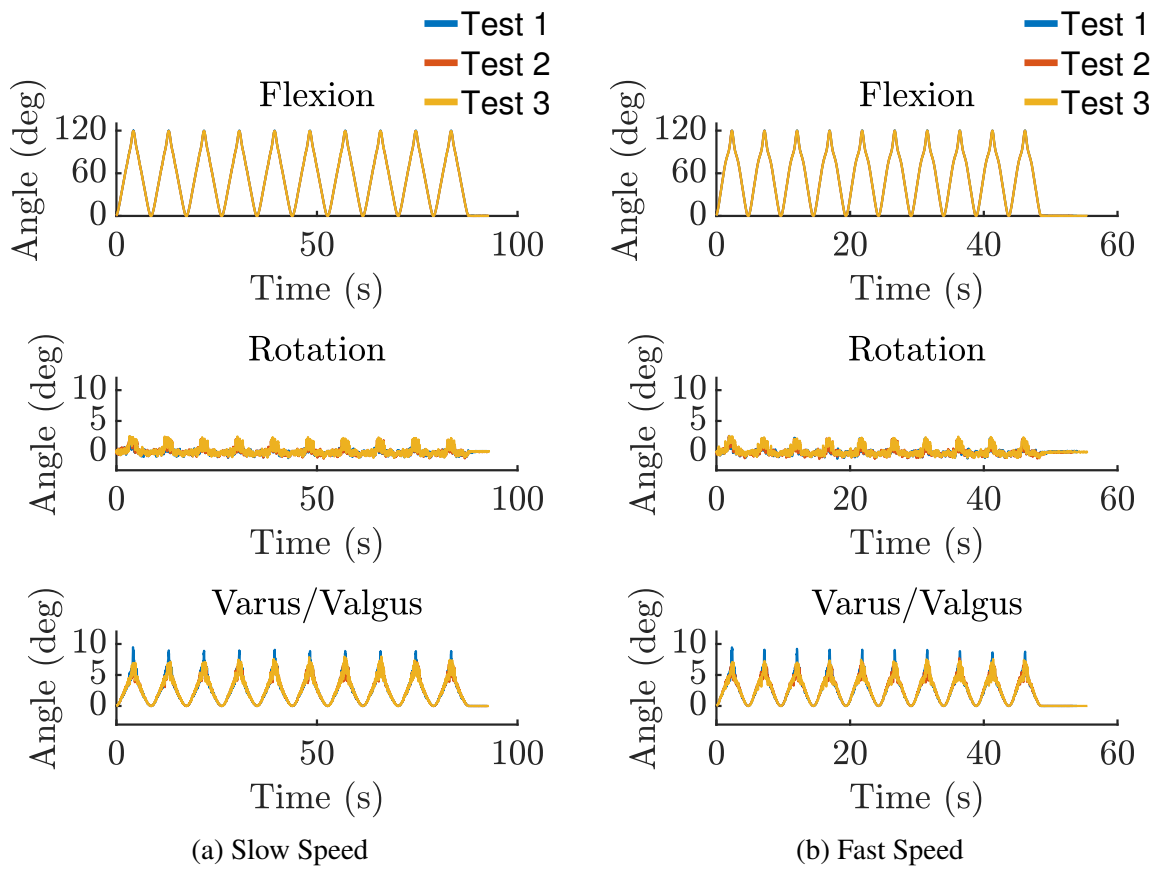
Figure 4.6: Angles computed during motion paths as the phantom limb travelled through 10 cycles using the gold standard motion capture system.

Table 4.1: Multi-Test Path Correlation

Slow	Lateral	Posterior	Gold Standard
Flexion	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00
Rotation	0.47 ± 0.34	0.68 ± 0.23	0.81 ± 0.08
Varus/Valgus	0.52 ± 0.31	0.92 ± 0.04	0.98 ± 0.01

Fast	Lateral	Posterior	Gold Standard
Flexion	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00
Rotation	0.45 ± 0.25	0.60 ± 0.35	0.83 ± 0.03
Varus/Valgus	0.67 ± 0.24	0.74 ± 0.14	0.98 ± 0.00

4.4.2 Path Amplitudes

Table 4.2 shows the mean sensor path amplitudes compared to the gold standard and their standard deviations computed with Equation (4.3). Flexion angles were shown to most accurately reach peak amplitudes in comparison to the gold standard for both trial speeds with a mean difference in degrees of (lateral-slow: 1.39, posterior-slow: 1.23, lateral-fast: 0.40, posterior-fast: 2.24). Amplitudes of rotation and varus/valgus paths have shown large standard deviations (ranging from 2.51 *deg* to 5.48 *deg*) which indicates that sensor-body alignment could introduce non-negligible error in these axes during patient usage.

Table 4.2: Multi-Test Mean Path Amplitudes

Slow	Lateral (deg)	Posterior (deg)	Gold Std. (deg)
Flexion	121.11 ± 2.92	118.49 ± 0.97	119.72 ± 0.33
Rotation	4.45 ± 5.48	4.04 ± 2.55	1.39 ± 0.76
Varus/Valgus	5.44 ± 2.63	4.92 ± 2.51	6.47 ± 0.43

Fast	Lateral (deg)	Posterior (deg)	Gold Std. (deg)
Flexion	120.39 ± 3.39	117.75 ± 1.92	119.99 ± 0.29
Rotation	4.41 ± 5.34	4.73 ± 3.23	2.14 ± 0.22
Varus/Valgus	5.63 ± 3.67	4.36 ± 2.75	7.70 ± 1.18

4.4.3 Root Mean Square Error

Table 4.3 shows the RMSE of each sensor test axis compared to the equivalent gold standard result calculated as shown in Equation (4.4). The mean error of flexion tests was shown to be

the largest (ranging from 3.29 *deg* to 3.52 *deg*), however, the range of motion in this axis was much larger. Varus/valgus showed low mean errors over three tests for both speeds (ranging from 1.42 *deg* to 2.30 *deg*). Mean errors of rotation varied more significantly between tests with a highest lateral error on Test 3 (4.39 *deg*) and a highest posterior error on Test 2 (4.03 *deg*) indicating rotation accuracy may be more influenced by placement when using the proposed rotation sequences.

Table 4.3: Root Mean Square Error to Gold Standard

Slow	Lateral (deg)			Posterior (deg)		
	Flex	Rot	Var	Flex	Rot	Var
Test 1	3.24	1.09	1.42	3.20	0.84	0.75
Test 2	3.66	1.80	2.08	3.68	2.71	1.49
Test 3	3.65	4.39	1.75	3.28	1.99	2.14
Mean	3.52	2.43	1.75	3.39	1.85	1.46

Fast	Lateral (deg)			Posterior (deg)		
	Flex	Rot	Var	Flex	Rot	Var
Test 1	4.43	1.37	2.75	4.02	0.97	0.78
Test 2	2.64	1.56	2.57	2.47	4.03	1.50
Test 3	3.41	4.24	1.58	3.38	2.49	1.98
Mean	3.50	2.39	2.30	3.29	2.50	1.42

4.4.4 Sensor Drift

The MARG sensors used in this study were in motion for $\approx 90s$ at slow and $\approx 50s$ at fast speeds without resting and in all tests there were only two occurrences of mean drift over 1 *deg* per ten-cycle test (slow-lat-var: 0.14 *deg/cycle*, fast-lat-var: 0.13 *deg/cycle*). It is expected that OA patients will pause intermittently within this time frame (allowing a gravity vector to be found) so sensor accelerometer correction is expected to help accuracy. Magnetic disturbances were assumed to be minimal during this experiment but error could accumulate in real-world situations due to non-uniform magnetic interference.

4.5   Discussion

Although the robot was programmed to flex through a linear path, non-negligible internal/external and varus/valgus rotation was observed in all tests due to the simulated motion of the artificial limb through flexion and extension. These additional motions are partly a result of the mannequin limb moving as a naturally constrained human leg, as well as soft tissue artifacts which are expected to be present during patient use (varying with leg size and muscle tone). Skin motion is a source of error in all external joint monitoring methods including gold standard gait systems and cannot be avoided using externally mounted sensors [106]. Despite some literature suggesting that additional motions can be ignored [61, 107, 108], knee motions such as varus/valgus thrust are important indicators in OA patients, indicating multi-dimensional measurement of the knee is significant [109].

Interpreting a quaternion as clinical angles is not trivial. Angle labels are assigned to rotations differently among literature, but they are often directly translated from three stages of independent local rotations constituting a traditional Euler representation [105, 107, 110–112]. There are many three-stage, non-unique, and correct decompositions that can be found which makes the extraction of clinical angles dependent on the sequence. Interpretation is further dependent on the chosen reference segment. For example, when the knee is flexed, internal/external rotation occurring about the lower segment with an upper limb reference is a change in varus/valgus of the upper segment with a lower reference. For this study, the authors chose to use the upper limb as a reference for clinical angle extraction. Since motion capture measurements were made with respect to a stationary frame aligned with the upper limb, the upper segment was chosen as a reference for sensor computations because it moved less than 3 *deg* in all axes during testing for more equal comparison.

A consequence of computing the change in rotation with respect to a reference segment is that accuracy will vary with alignment of the reference sensor to its body frame. An assumption of this system suggests that one of the sensors (reference) can be placed on a patient with correct alignment to the limb frame. Visual alignment has been previously tested with less than 5 *deg* error in a different wearable system [111]. Since the change of rotation is with respect to the reference frame, a misalignment of the non-reference sensor should not introduce

error (kinematic crosstalk [113]) associated with knee motion occurring around a misaligned reference. Based on testing results, it is expected that poor alignment of the reference sensor will introduce this error. The authors suggest that correct placement is aided by the custom sensor case design used, which is elongated along the limb length, concave to conform and deter twisting, and held in-place using stretchable straps to prevent sensor liftoff. Further testing should take place concerning the correct attachment of a reference sensor to patients for improvement of rotation accuracy. Case design could be improved to accommodate wider straps to prevent the twisting of sensors against the skin. It can be seen in Figure 4.1b that the straps used are more narrow than the 3D-printed slots on the red sensor cases. It is expected that at some placements, the sensors were able to rotate slightly independent of the straps during placement which should be avoided. Sensor fixation could be improved during patient usage by adding medical grade two-sided adhesive tape under the hook and latch straps.

In addition, further experimentation should consider an anterior sensor placement on patients for two reasons: the anterior lower limb could provide the better reference placement in patients since there is typically less soft tissue on the front of the shin, and this placement option would prevent patients from sitting on the sensors during activity.

The current work explored the reliability of extracting joint angles with the placement and replacement of sensors on a phantom limb. As an area of extension for future work, further analysis could focus on deliberate misplacement of sensors to examine the measurement error. If data are collected in a repeatable fashion, sensors could possibly be virtually rotated about their local frame in small increments to further understand the system sensitivity. A small virtual rotation about a sensor's z frame would simulate placement errors in the varus/valgus measurement while local rotations about a sensor's local x frame would simulate misalignment in the internal/external rotation parameters. Since skin motion is problematic with externally mounted hardware, skin motion could also be introduced artificially by incorporating non-constant rotation augmentations about local sensor frames while examining measurement accuracy.

Literature indicates that the standard error of measurement (SEM) in post-TKR patient flexion measured manually by experts is 4.1 *deg* while patient flexion has been shown to fluctuate up to 9.6 *deg* between measurements so system errors ≈ 4 *deg* can be declared clinically

insignificant [114]. Wearable systems implemented using goniometers have been shown to produce $\approx 3 - 4$ *deg* of error during usage and previous inertial systems have reported RMSE of $\approx 2 - 4$ *deg* [45]. The experiment performed has shown that the three clinical angles can be extracted within these ranges of error.

Joint velocities of OA patients monitored may exceed the speeds tested in this study. The method of joint angle extraction was tested primarily, and similar accuracy at higher joint velocities will depend on the inertial sensors used, which have previously been shown capable of operating at more realistic velocities [45, 68].

4.6 Contributions

The proposed system using independent attitude estimations meets many of the desired qualities reported to be desirable in wearable systems for patients with OA, such as unobtrusiveness, small size, and low cost [87]. Minimal calibration facilitates usage for patients with limited mobility, making this tool appropriate for home, ambulatory, or clinical use as a replacement for more expensive instrumentation. All clinical angles have shown a low RMSE in degrees for both placements (flex: ≤ 3.52, rot: ≤ 2.50, var: ≤ 2.30) and low degree difference in mean cycle peaks (flex: 1.32, rot: 3.71, var: 2.00) compared to a gold standard, however, it has been mentioned that poor visual placement of the sensors could introduce larger error in practical use.

From these data, joint velocities and accelerations can be computed as well, potentially allowing for discovery of important knee parameters during patient usage. This proposed method of computing joint angles using absolute attitude estimations can be used in any body orientation free of singularities, allowing knee joint monitoring through any scripted or unscripted activity. The proposal of this system and validation has been published in *IEEE Transactions on Biomedical Engineering* [115].

Chapter 5

Deriving Knee Performance Measures from the Timed-up-and-go Test

5.1 Introduction

A valid tool is important for functionally evaluating patients, however, an appropriate protocol for instrumentation is crucial for repeatable measurement across repeated visits or patient populations. As previously detailed in Chapter 3, scripted performance tests offer the ability for direct comparison instead of performing independent observations on subjects while they complete different tasks. The timed-up-and-go (TUG) test is a favourable test to evaluate knee function since it combines several joint stressing activities and requires no additional equipment to complete. This chapter details the implementation of a strategy to extract relevant knee joint parameters from instrumented timed-up-and-go tests that can be used to evaluate patient function post-surgery.

The contents of this chapter begin with a preliminary experiment to ensure the practicality of the instrumentation strategy and if it will be effective when deployed in the clinic. Patients were instrumented in the clinic by Megan Fennema as a method of functional evaluation for her work assessing patient outcomes. All information was de-identified before being used for the analyses presented in this chapter.

Inter-patient functional differences are examined in healthy subjects completing instrumented TUG tests. The remainder of the chapter focuses on extracting more granular knee-

specific joint parameters that can be used to differentiate positive and negative functional outcomes in knee replacment populations. Extracted metrics are validated through correlating to self-reported outcomes at one-year post-operation.

5.2 Instrumentation Validity for Timed-up-and-go test

Before investigating the timed-up-and-go in total knee replacement patients, a primary experiment was used to determine the suitability of the developed instrumentation methods for the TUG test. This primary investigation had two objectives: 1) Subjectively ensure appropriate flow for deployment in the orthopaedic clinic, and 2) Verify that the minimal instrumentation strategy detailed in Chapter 4 is sufficient to measure functional differences in subjects executing the TUG test.

For this initial experiment, high precision data logging inertial sensors were used (YEI 3-Space Data Logger, Yost Labs) in place of the Mbientlab sensors used in Chapter 4. Since the knee instrumentation strategy depends on raw orientation outputs, any sensor capable of outputting attitude estimations can be used. These sensors were readily available and advertised highly accurate estimations (± 2 deg error in all axes). The internal sensor fusion algorithm used a Kalman filter to integrate the raw gyroscope, magnetometer, and accelerometer data into usable attitude estimations. A large limitation of these sensors was their wired interface: they were capable of logging timestamped data but these logs must be transferred over a serial cable connection. A custom Matlab script was created to enable sensor streaming and perform data analysis. A polling technique was used to obtain soft real-time synchronization at approximately 25 Hz. Once sensors were connected to a laptop running the Matlab script, sensors were affixed to subjects and a 30 second recording session. This limitation was overcome by using a very long USB cable which permitted subjects to hold the cable elevated in the air while they completed the tests and were able to turn without difficulty. Unfortunately due to the limitations of the tethered cables, only a single knee was instrumented at a time but the methods below have been extended to both knees, which will be applied further in the chapter for knee replacement patients. For each healthy participant, five TUG trials were recorded.

5.3 Performance Test Segmentation

As discussed in Chapter 3, the total test completion time is a valid measure that has been used to evaluate impairments and even predict falls in elderly populations. However, sensor instrumentation can be used to extract more granular measures. Since total completion time has been proven relevant to outcomes following TKR, it was expected that test segmentation into all sub-activities could also reveal important information [116].

Since sensor data were recorded as raw quaternion orientation information relative to a common global frame, raw quaternion rotation thresholds could be used to indicate segmentation points. Where more context was required, left and right knee flexion angles were used. To derive a series of joint flexion angles for the duration of the test, matched timestamped unit quaternions of the form

$$q = [w\,x\,y\,z]$$

where $w, x, y, z \in \mathbb{R}$ and $0 \leq w, x, y, z \leq 1$, with unit magnitude as:

$$|q| = \sqrt{w^2 + x^2 + y^2 + z^2} = 1 \tag{5.1}$$

from the upper and lower leg were used to compute a difference in rotation q_Δ between the two segments as:

$$q_\Delta = q_{ref}^{-1} * q_{rel} \tag{5.2}$$

similarly to the range of motion calculations of the hip and knee *Nguyen et al.* presented [78]. This q_Δ represents a minimal difference in rotation between the two upper and lower sensors. Knee flexion was further extracted as the rotation about the primary axis using the methods presented in our previous sensor system validation work [88]. Similarly, the lower shin sensor was used as a reference due to reduce skin motion compared to the upper thigh mounting location. An overview of segmented test sub-activities alongside series of left and right flexion angles can be seen in Figure 5.1.

To provide more detailed metrics from a recorded TUG test, the total recording was seg-
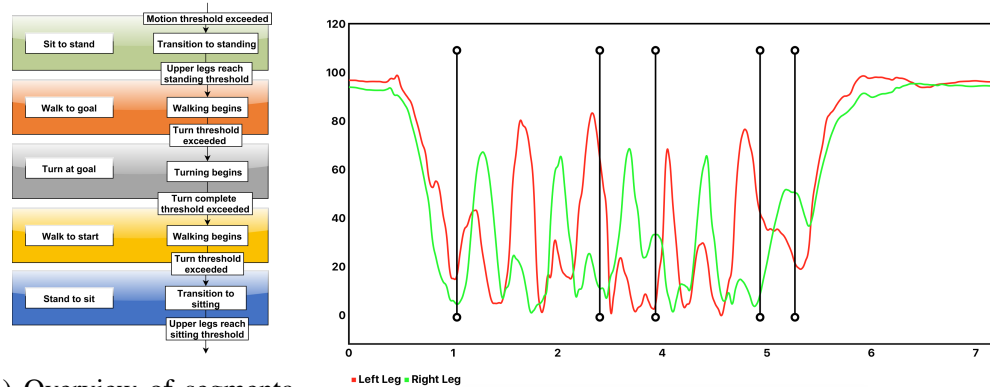
(a) Overview of segmentation process.

(b) Recorded test segmented with a Matlab script.

Figure 5.1: TUG segmentation and left/right knee flexion angles (degrees) from a recorded test.

mented into five segments: sit to stand, walk to goal, turn at goal, walk to start, and stand to sit. It is expected that each of these segments should be analyzed independently as they roughly are indicative of each combined activity evaluated in the test. The proposed segmentation algorithm relies on the detection of key test indices (Table 5.1) that separate the activities of standing, walking, turning, and returning to sit in the starting position.

Test Recording Trimming

Before segmentation of the TUG test, the total data recording is first trimmed to remove observer influence. It was noted that on the "go" command, some participants hesitated before beginning to move and others began the test early even though the instructions were explained clearly (especially on the first trial) and the beginning could be missed. Existing work has used a change in pitch of the torso for a test start detection however from observation it was seen that some participants moved forward and backward in their chair before the test start and this should not indicate the beginning of a test in our population even if instrumentation extended to the torso.

Beginning from the first recorded sample, the total test recording is trimmed to the first motionless state which prevents unintentional motion at the start of the test that could trigger a

Table 5.1: Summary of segmentation indices used to separate the test into sub-activities for further analysis and metric extraction.

Index Name	Description
recordingStartIndex	Trim the recording to the point where the subject's legs are still. Removes false test starts if the recording is started before the subject is motionless.
testStartIndex	The start of the TUG test as determined by the first signs of leg motion. Indicates the subject is preparing to stand.
chairLiftoffIndex	The subject's thighs have exceeded a vertical threshold indicating they have lifted off the chair at the start of a stand action.
standEndIndex	A large change in extension of the left or right knee is used to detect the end of the standing activity.
turnStartIndex	A mean index between the left and right legs reaching a rotation about the world vertical axis indicates the start of a turn.
turnEndIndex	A second mean index of the time when both legs are nearly facing the starting chair position indicates the end of a turn.
sitStartIndex	The sit start is detected when the subject finishes turning to face the direction of the floor-marked goal.
testEndIndex	The sitting task is expected to fill the remainder of the test and the end is detected when either leg is flexed and mostly still. This accounts for subjects who sit with a single leg extended in front of them.

false start. For this task, a change in knee flexion f_Δ at time t is computed as

$$f_\Delta = f_{t+1} - f_t \tag{5.3}$$

and the recording start is first trimmed so f_Δ of both left and right knees is less than 0.25 deg/t where t is also the sampling period of the sensor system. This removes any unnecessary movement from the subject if the recording is begun before becoming stationary in the chair. The *recordingStartIndex* is determined to be the first instance with no motion after recording is started by the tester.

Detecting the TUG Start

Following the detection of a motionless state that marks the beginning of a valid recording, the next step is to detect when a patient actively intends to begin their test. While *Higashi et al.* showed a single thigh sensor exceeding a motion threshold about the pitch axis could determine this start index, it was observed that some of our subjects exhibited a "preloading effect" where they flexed one or both knees in anticipation of standing while others rose straight to a stand more fluidly. The first indication of motion was determined by observing flexion angles and for the current work an absolute change of only $3deg$ over a period of 0.25s was sufficient to detect the first signs of motion before a stand activity. This f_Δ was computed over five samples to reduce the likelihood of a false trigger due to low-magnitude sensor noise. This low motion threshold and other values expressed in the remaining segmentation steps were chosen based on observation of animated avatars recreated by raw sensor readings using a leg mesh model created with Blender 2.8. Since there is no generally accepted threshold for the TUG segmentation it is expected that consistency across subjects is of most importance. An example of these avatars used to detect motion thresholds can be seen in Figure 5.2.

Sit-to-Stand

After detecting the test start as the first registration of motion, the standing activity is the only possible transition from sitting to walking. This was chosen to begin at chair liftoff, which was found by computing the quaternion rotations of each upper leg sensor relative to their

(a) Subject at test start in seated position.

(b) Subject nearing the end of a stand.

(c) Subject taking the first step after standing.

(d) Taking a full step during the walking segment.
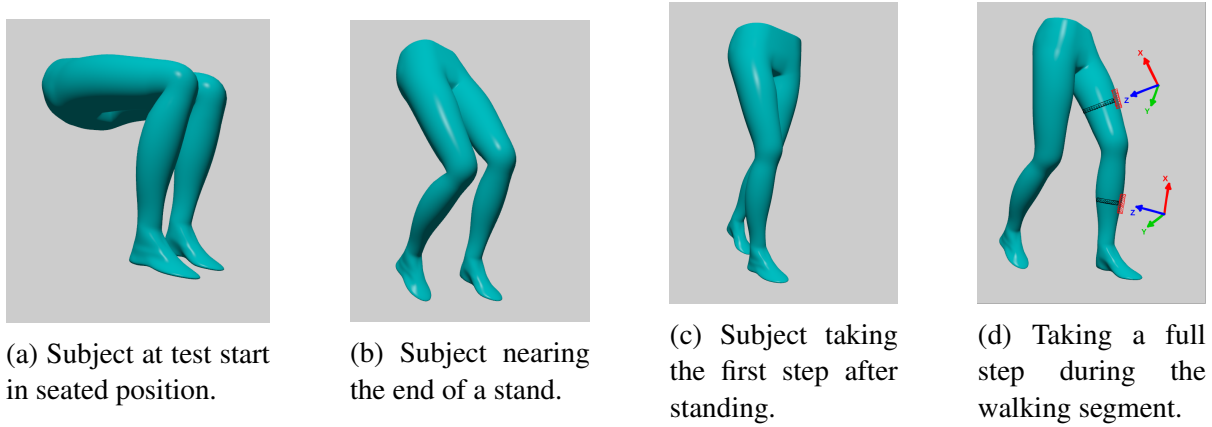
Figure 5.2: Animated avatar recreating a subject's TUG test based on recorded sensor readings. Ankle flexion is held at a right angle since no foot motion is recorded by the system. Figure 5.2d contains sketches indicating sensor placements and local coordinate frames.

measure at the test start index. By re-purposing Equation (5.2), we can find a q_Δ independently for left and right upper sensors with q_{ref} being the past reading at the test start index and q_{rel} as the present measure. Due to the order of multiplication in Equation (5.2), q_Δ is the rotational difference between past and future locations of the same sensor measured about the sensor local frame at the test start index. Because the upper sensor is on the top of the thigh, the y axis of the local frame is of interest to detect liftoff (Figure 5.2d). Furthermore, it would be expected that skin and clothing motion will introduce rotation in the x and z axes of the local frame. Using the unit quaternion property in Equation (5.1), nullifying the x and z components of q_Δ will emulate a rotation needed to bridge the past and present sensor locations as if no rotation took place in those axes. With a zero value for x and z, q_Δ can be normalized using Equation (5.1) to get a compensated unit rotation. It was found that a value of 0.05 in the y component of this compensated difference was sufficient to detect chair liftoff. Through an extension of Euler's formula:

$$q = e^{\frac{\theta}{2}(u_x i + u_y j + u_z k)} \tag{5.4}$$
$$= cos\frac{\theta}{2} + \left(u_x i + u_y j + u_z k\right) sin\frac{\theta}{2}$$

the angle θ about any of the imaginary quaternion axes can be found as:

$$\theta = 2 \times \arcsin(u) \tag{5.5}$$

It can be seen that this threshold corresponds to 0.10 *rad* or approximately 6 *deg* of rotation about the *y* axis.

The end of the stand was detected when this same *y* value exceeds a magnitude of 0.6 (approximately 75 *deg* using Equation (5.5)) on either left or right legs. It should be noted that often times the leg taking the first step does not meet this threshold as subjects elevate into their first step with the dominant knee remaining flexed into the first step.

Turning

In unscripted tasks, turning is a difficult activity to detect. During daily motion, many small turns are executed as subjects navigate their environment and these cannot be distinguished from larger turns. During the TUG, it is known that patients must execute a 180 *deg* turn at the end of the walking phase. Since the test is scripted, turn detection is simplified. Starting from the stand end index found in the previous step, the turn start index is found similarly to the chair liftoff index by computing a series of q_Δ and nullifying unwanted components for the left and right legs but instead of the *y* axis of the upper sensor, the *x* axis of the lower sensor is used representing the vertical world axis of the lower leg frame. A turn marker for each leg is set when the absolute value of *x* reaches 0.1. Once both left and right leg turn markers are set, the turn start index is found to be a mean of both markers. The turn direction is found by examining the sign of the change. Due to the left handed coordinate frame of the sensors used and the positive *x* axis pointing opposite of the floor, the "left hand rule" dictates a negative value is turning left. The end of the turn is detected similarly with the magnitude of the compensated *x* value being 0.9.

Walking

Walking phases of the test were simply extracted as the filling sections between detected sit-to-stand and turning activities as well as the end of turning at goal to turning to a seated end

position. This strategy is effective since turns can be identified confidently and walking is the only activity patients will be completing during these sections.

Stand-to-Sit

The sitting activity is challenging to segment since it was observed that the start of a turn was often indistinguishable from a sit. Many subjects turned while sitting and others first completed their turn entirely. It is expected that a combined turn and sit indicates more confidence in sitting ability than an individual who must turn and align themselves with the chair before starting an independent sit. For the current work, a hybrid strategy was used to group these together. The start was detected as the start of a second turn using the same criteria as the previous turn detection. The sit activity was expected to take the remaining time and was ended on the test end detection. This was found to be when both left and right flexion angles were greater than 45 *deg* and both legs returned to a motionless state.

Segmentation Evaluation

Due to difficulties establishing ground truths to test the accuracy of automatic segmentation thresholds and the lack of accepted sub-activity definitions, the current work has not used external human evaluators to confirm segmented indices [80]. Since tests were recorded as absolute attitude estimations of each of the four leg segments, animations using a leg mesh model (Figure 5.2) were used to recreate patient tests and verify segmentation was successful.

5.4 Healthy Participant Inter-subject Functional Differences

During test recording of the healthy participants it was observed that the sit-to-stand and turning activities seemed to vary the most between individuals. It was also noted that the number of steps and overall range of knee flexion would vary between participants. Unfortunately for this study subject leg-length and height was not recorded. It was also previously noted that some subjects demonstrated a preloading effect where knee flexion was increased from the natural sitting position before chair liftoff in an attempt to build momentum for the stand. Due

to these observations, a comparison of these noted differences was performed between all participants. Since each subject executed the test in five trials, the inter-subject deviation could be computed and compared to the intra-subject deviation to hint at the ability to detect functional differences between subjects. For comparison, sit-to-stand and turning completion times were computed as the time difference between chairLiftoff-standEnd and turnStart-turnEnd indices, respectively (Table 5.1). The number of steps was counted manually by examining animated avatars of subject tests and recording complete cycles of heel strike, mid-stance, and heel liftoff cycles. Preload flexion was computed as a sum of wavelength (or the total absolute length of signal) of flexion angles from the testStart to the chairLiftoff index. The percentage difference was computed between inter-subject and intra-subject differences for a rough indication of functional differences between subjects. The following computation was used for percentage difference:

$$\% \text{ Difference} = 100 \times \frac{\|\text{Inter-subject Mean} - \text{Intra-subject Mean}\|}{(\text{Inter-subject Mean} + \text{Intra-subject Mean})/2} \tag{5.6}$$

5.4.1 Results

The segmentation was successful in segmenting all 28 trials for all subjects (2/30 did not record correctly due to loss of communication with the sensors) confirmed with visual inspection of the angular data recorded.

Time taken during sit-to-stand activities was found to vary on average $\pm 0.23s$ per person and $\pm 0.36s$ between subjects (Table 5.2). Time to rotate around the goal was found to be on average $1.2s$ with $\pm 0.3s$ deviation between subjects. The average number of steps taken with the instrumented leg was found to be 5.8, varying ± 0.4 steps between subjects (Table 5.3). Maximum range of motion during walking stages was found to be $62.9\,\text{deg}$ deviating on average $\pm 4\,\text{deg}$ per person and $\pm 6\,\text{deg}$ between subjects. On 5 of 6 subjects, an observable preloading effect was captured during initiation of the sit-to-stand stage with a mean flexion angle increase of $2deg \pm 1$ deg between subjects with the observed effect.

Table 5.2: Inter vs. intra subject differences for the derived metrics

Metric	Inter-subject deviation	Average intra-subject deviation	Percent diff.
Sit to stand time (s)	0.36	0.23	44.07
Turning time (s)	0.28	0.24	15.38
Preload flexion (deg)	1.15	1.51	27.07
Steps taken	0.39	1.37	111.36
Range of motion (deg)	5.95	4.02	38.72

Table 5.3: Values of primitive metrics derived from the timed-up-and-go test from a set of six healthy participants.

Subject	Sit to stand (s)	Turning (s)	Preload flexion	Mean steps taken	Range of motion
1	1.57 ± 0.47	1.07 ± 0.12	1.20 ± 0.99	5.80 ± 0.75	67.84 ± 3.03
2	1.92 ± 0.14	1.33 ± 0.40	1.63 ± 1.53	5.80 ± 0.98	69.85 ± 1.57
3	1.77 ± 0.29	1.45 ± 0.29	3.09 ± 2.63	5.40 ± 0.49	53.71 ± 2.98
4	1.11 ± 0.26	0.68 ± 0.05	3.10 ± 1.90	5.20 ± 5.20	57.82 ± 3.25
5	1.24 ± 0.15	1.41 ± 0.15	0.73 ± 0.53	6.00 ± 0.00	67.69 ± 1.13
6	0.94 ± 0.09	1.48 ± 0.45	0.00 ± 0.00	6.40 ± 0.80	60.34 ± 12.15

5.4.2 Discussion

Analysis of time taken during the sit-to-stand stage and range of motion shows there was a 44% and 39% difference in deviation between test subjects over deviation between sequential trials from the same subject. Large intra-subject deviation was observed for the number of steps however Subject 4 could be considered an outlier. After removal, a more expected deviation of 0.60 exists for the five remaining subjects. This suggests measurable performance differences between subjects exist despite a healthy sample. Additionally, notable differences are present in range of motion between subjects indicating an individual difference in joint usage while walking. The presence (or lack of) the observed preloading feature could indicate a physiological characteristic or strategy for standing. This initial test has used very few participants so no further statistical analysis has been performed. Intra-subject deviation did not seem to observably increase with the number of tests performed sequentially by the same subject, which could have indicated fatigue.

The results of this experiment showed that the absolute body orientation data were reliable for segmenting activities on scripted tests, and that differences in individual test execution could be discriminated despite a healthy sample of volunteers. Several limitations of this study

were identified however, such as testing requiring subjects to be connected to a wired laptop. Additionally, testing was performed on a small sample of healthy individuals, whereas the target demographic was intended to be patients with knee OA. A major component missing from this testing was confirming with a ground truth measurement technique that the angles being measured were sufficiently accurate to provide useful measurements of the knee. It was noted that for testing to take place safely in the clinic with knee replacement patients, lightweight and unobtrusive sensors would be mandatory. The wired sensors used in this small study were not practical and all future work will proceed with the portable wireless sensor packages described and used in Chapter 4.

The results of this experiment were presented in a talk at the 2017 Imaging Network Ontario Symposium under the development of novel therapies for bone and joint diseases category.

5.5 Granular Knee Evaluation Parameters

Although the use of this sensor instrumentation system has allowed for segment time extraction free of human introduced error during test segmentation, the goal of observing functional performance of the knee joint has not been entirely fulfilled. Test completion times still do not offer granular information as to how the test was completed or provide any measure of joint function. Patients may be able to quickly navigate the test by favouring an uninjured leg or by other compensation. This section details the extraction of more specific joint metrics, in addition to completion times, by category.

A total of 55 metrics have been used to evaluate patient function in the current work. Since left and right labels for metrics are not helpful to compare patient populations and the operative side was known, any left and right differentiation was replaced with operative (op) or non-operative (non-op) prefaces as applicable.

5.5.1 Temporal

Firstly, the completion time for each segment of the TUG was found in addition to the total test completion time using the temporal difference between each of the transition indices

determined. These six times include: total completion, sit to stand, walking to the goal, turning about the goal, walking to the chair, and stand to sit (Table 5.4).

Table 5.4: Descriptions of temporal metrics extracted from TUG test segmentation indices.

Total Time	The total time taken to complete the test from the test start index to the test end index
Sit To Stand	Duration beginning from the chair liftoff index to the stand end index
Walking To Goal	Time spent walking from the stand end index to the beginning of the turn at the turn start index
Turning At Goal	Duration of turning action from turn start index to turn end index
Walking To Chair	All time spent from the turn end index to the sit start index
Stand To Sit	The remainder of the test starting from the sit start index to the test end index

5.5.2 Sitting Position

The instantaneous knee flexion of both left and right knees is recorded at the test start index to obtain a measure of sitting position. This is also measured at the test end index. Additionally the asymmetry of the left and right knees at this point is found to be the absolute difference in flexion between the two legs (Table 5.5).

Table 5.5: Descriptions of instantaneous sitting flexion metrics from the test start and end indices.

Operative Start Flexion	Operative leg instantaneous flexion recorded at the test start index
Non-operative Start Flexion	Non-operative leg instantaneous flexion recorded at the test start index
Operative End Flexion	Operative leg instantaneous flexion recorded at the test end index
Non-operative End Flexion	Non-operative leg instantaneous flexion recorded at the test end index
Start Flexion Asymmetry	Difference in starting flexion angle betweeen the operative and non-operative legs
End Flexion Asymmetry	Difference in ending flexion angle betweeen the operative and non-operative legs

5.5.3 Total Accumulated Motion

For each of the four leg-mounted sensors, a total accumulated motion metric was computed to obtain a measure of total leg segment motion (Table 5.6). For each sensor s, this measure is found by computing a series of q_Δ iteratively from the test start index to the end of the test and summing the angle extracted from the scalar w component of all q_Δ rotations. It can be expressed as:

$$s_{total\,motion} = \sum_{n=testStartIndex}^{testEndIndex-1} |2 \times \arccos(\langle w_n \rangle)|, \tag{5.7}$$

$$\langle w_n, x_n, y_n, z_n \rangle = q_{\Delta_n} = q_n^{-1} * q_{n+1}$$

Table 5.6: Descriptions of total accumulated sensor motion.

Operative Accumulated Thigh	Summation of the overall change in angle of the operative upper sensor for the entire test
Non-operative Accumulated Thigh	Summation of the overall change in angle of the non- operative upper sensor for the entire test
Operative Accumulated Shin	Summation of the overall change in angle of the operative lower sensor for the entire test
Non-operative Accumulated Shin	Summation of the overall change in angle of the non-operative lower sensor for the entire test

5.5.4 Accumulated Clinical Angles

Additional accumulated parameters were extracted with clinical interpretation in mind to approximate anatomical angles used to describe the knee [53]. Series of medial/lateral rotation, and varus/valgus offset angles were computed for each leg in addition to the flexion angles computed in Section 5.3 with f_Δ computed in Equation (5.3). The total accumulated flexion metric for a single leg is computed in Equation (5.8) but can be extended to the other two anatomical degrees of freedom (Table 5.7).

$$\sum_{n=testStartIndex}^{testEndIndex-1} |f_{\Delta_n}| \tag{5.8}$$

Table 5.7: Descriptions of accumulated extracted clinical angles.

Operative Accumulated Flexion	Summation of extracted operative flexion/extension changes in the primary knee axis from the test start index to the test end index
Operative Accumulated Rotation	Summation of extracted operative internal/external rotation changes in the secondary knee axis from the test start index to the test end index
Operative Accumulated Varus	Summation of extracted operative varus/valgus changes in the tertiary knee axis from the test start index to the test end index
Non-operative Accumulated Flexion	Summation of extracted non-operative flexion/extension changes in the primary knee axis from the test start index to the test end index
Non-operative Accumulated Rotation	Summation of extracted non-operative internal/external rotation changes in the secondary knee axis from the test start index to the test end index
Non-operative Accumulated Varus	Summation of extracted non-operative varus/valgus changes in the tertiary knee axis from the test start index to the test end index

5.5.5 Step Detection and Evaluation (33)

Gait analysis has been used to evaluate function on many occasions with success. With the current instrumentation system, traditional gait parameters are difficult to extract without further information such as leg length and the sensor distance from the knee joint. Since this information was not collected from the study population, a primitive step detection method was implemented and more specific stepping parameters were extracted from detected steps.

A step template was first generated by manually selecting series of step flexion angles from subject flexion charts generated from the TUG test (Figure 5.3) using custom developed software. Subjects from initial testing were first used but the created template was updated with total knee replacement patient data after obtaining data from subjects undergoing total knee replacement. Steps were selected based on flexion angles starting at heel strike and moving through mid-stance to toe-off and all steps were verified using the animated leg mesh to ensure validity. The step template was trimmed to the mean length of all selected steps to obtain an average step pattern. This mean step pattern was then correlated over all segmented walking portions of patient tests and large responses of 0.8 or greater indicated a step.
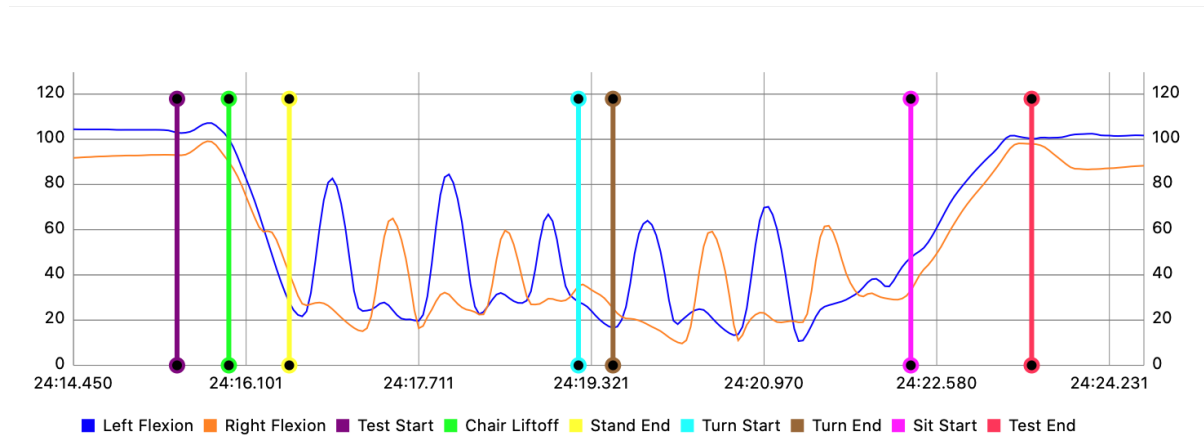
Figure 5.3: Left and right knee flexion/extension angles (degrees) during the TUG test with segmentation indices using the specified motion thresholds.

For each step, 11 metrics were computed for each left and right leg. The minimum and maximum flexion angle observed was recorded and used to determine a flexion range as well. The maximum and minimum flexion velocities were obtained using delta changes in the step flexion series and likewise, maximum and minimum flexion acceleration were obtained as a delta change in flexion velocity. Similar to previous accumulated motion metrics, this was obtained for the upper and lower sensor independently for only a detected step instead of the entire test. Since the quality of steps varies during the short walking portions of the TUG as patients rise from a stand and complete their turn, mean values for all step metrics were computed by averaging each step metric over all detected steps. An asymmetry value was computed using the absolute difference for each of the 11 metrics bringing the total stepping parameters to 33 (Table 5.8).

5.6 Metric Validation with Total Knee Replacement Patients

The remainder of this chapter will employ similar methods of automatic segmentation of the TUG test using the minimal wearable system detailed in Chapter 4. The goal will be to determine if the proposed metrics can differentiate positive and negative knee replacement outcomes as measured using the current standard of care evaluation techniques (self-reported questionnaires). It is expected that these relevant parameters can be used to provide individu-alized evaluations for patients undergoing total knee replacement in hopes that early problems

Table 5.8: Descriptions of step metrics for a single leg. All metrics are computed for both operative and non-operative sides and asymmetries of each metric are found to be the absolute difference in op/non-operative values.

Step Maximum Flexion	A mean value of the maximum observed flexion angle during a detected step, averaged over all steps.
Step Maximum Flexion Range	A mean value of the observed flexion range during a detected step, averaged over all steps.
Step Maximum Flexion Velocity	A mean value of the maximum observed flexion velocity during a detected step, averaged over all steps.
Step Maximum Extension Velocity	A mean value of the maximum observed extension velocity during a detected step, averaged over all steps.
Step Maximum Flexion Acceleration	A mean value of the maximum observed flexion acceleration during a detected step, averaged over all steps.
Step Maximum Extension Acceleration	A mean value of the maximum observed extension acceleration during a detected step, averaged over all steps.
Step Accumulated Thigh	A mean value of the summations of the overall change in angle of the upper sensor for the duration of a step, averaged over all steps
Step Accumulated Shin	A mean value of the summations of the overall change in angle of the lower sensor for the duration of a step, averaged over all steps
Step Accumulated Flexion	A mean value of the summations of extracted flexion/extension changes over the duration of a detected step, averaged over all steps
Step Accumulated Rotation	A mean value of the summations of extracted internal/external rotation changes over the duration of a detected step, averaged over all steps
Step Accumulated Varus	A mean value of the summations of extracted varus/valgus offset changes over the duration of a detected step, averaged over all steps

can be detected for further intervention to help improve overall outcomes.

5.6.1   Data Collection

To evaluate the effectiveness of extracted parameters in differentiating patients with positive and negative recovery outcomes, IRB approval was obtained to recruit and instrument primary TKR patients coming into the clinic for their one or two year follow-up appointment

(Appendix A). Individuals were excluded if they had language or cognitive barriers, a neuro-muscular disease, a TKR on the contralateral leg within one year of participating, or underwent a revision surgery on their primary TKR. Each patient was instructed to complete three trials of the TUG test in the clinic using a standard hospital chair and the sensor system. There was no further test standardization and patients completed tests wearing their choice of footwear. Subjects began seated in a standard exam room chair with both feet planted on the floor while sensors were attached by an observer and the measurement application was initialized. They were then instructed to begin the test on the command "go", walk to a goal three metres away (marked on the floor using tape), turn about the goal, walk back to the starting position and sit back down with their back against the rest. Subjects were also instructed to complete the test safely and to move at a comfortable but swift pace.

In addition to sensor instrumentation, patients completed a series of standard PROMs as a gold standard measure of surgical success. These included the: Short Form 12 (SF-12), Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC), Knee Society Score (KSS), and University of California Los Angeles (UCLA) activity score questionnaires. Patient satisfaction was extracted as sub-components of the KSS survey with the satisfied group corresponding to approximated answers of "very satisfied" or "satisfied" and dissatisfied patients labelled as those responding "neutral", "dissatisfied", or "very dissatisfied".

5.6.2 Methods

All metrics detailed in section Section 5.5 were computed for all tests executed by patients. Independent unpaired t-tests were computed on each metric to identify significant differences between both satisfaction groups to highlight their effectiveness to distinguish satisfied and dissatisfied patients. Multiple t-tests were chosen over one-way ANOVA to observe differences in each metric independently opposed to considering all group differences together. In addition, Pearson correlation coefficients were computed for all metrics to examine the strength of their linear relation to patient satisfaction, which was evaluated using the Knee Society Score satisfaction sub-score. In the current work correlations were identified as very strong (> 0.80), moderate (> 0.60), fair (> 0.3), and weak (> 0.20) [117]. Correlations with less strength are

not considered notable.

Principal component analysis (PCA) was performed on extracted functional metrics to examine the feature variance between patients. All features were first standardized to have zero mean and unit standard deviation across the entire sample.

5.6.3 Results

Recruited patients (n=82) were instrumented on the day of their clinical appointment and performed the TUG test in the orthopaedic clinic hall. Despite participating in the functional testing, some patients (n=10) did not correctly complete their PROMs and omitted their satisfaction score. The remaining study population was mostly satisfied with 17% self-reporting dissatisfaction according to the study separation criteria. Patients that did not complete their satisfaction score measure were excluded from future analysis.

Dissatisfied patients were younger (57 vs. 68 years) and had a higher mean BMI (36 vs. 31 kg/m^2). All individual PROM measures recorded were also found to be significantly different between satisfaction groups (Table 5.9).

Table 5.9: Mean ± standard deviation of PROMs and demographics between satisfied and dissatisfied patients.

Mean ± Std. Dev.	Satisfied	Dissatisfied	P
Age (yrs)	68.18 ± 9.25	57.00 ± 9.55	<.001
BMI (kg/m2)	31.23 ± 6.04	36.42 ± 10.00	0.019
UCLA Activity Score	6.07 ± 1.45	4.25 ± 1.96	<.001
SF-12 Mental	55.98 ± 7.47	45.93 ± 12.65	<.001
SF-12 Physical	43.49 ± 9.66	32.17 ± 8.61	<.001
WOMAC Pain	82.2 ± 16.38	53.75 ± 22.68	<.001
WOMAC Stiffness	72.03 ± 16.79	43.75 ± 30.39	<.001
WOMAC Function	78.94 ± 16.47	50.98 ± 21.33	<.001
WOMAC Total	78.86 ± 15.02	50.62 ± 19.71	<.001
KSS Symptoms	22.42 ± 3.80	17.58 ± 4.48	<.001
KSS Expectations	10.10 ± 2.83	4.91 ± 1.64	<.001
KSS Functional Activities	72.38 ± 16.07	41.88 ± 19.42	<.001

The step detection template was updated to reflect steps taken by patients following TKR. Still using the custom developed software, 135 left and right step patterns were manually selected from random recorded patient tests. The updated template was used in the generation of

the following step metrics.

After computing independent unpaired t-tests it was found that 17 of the developed metrics were significantly different between satisfaction groups (Table 5.10). The only metric with a moderate correlation to satisfaction was the total test time (-0.60). All additional segment times show fair correlations except the time taken to turn about the goal, which was only weakly correlated (-0.26) (Table 5.10).

Table 5.10: All metrics that have at least a fair correlation (> 0.20) or were significantly different between satisfied and dissatisfied patient groups. Negative correlations indicate lower metric values are more related to positive satisfaction scores.

Metric Name	Metric Description	Satisfied Mean ± Std. Dev.	Dissatisfied Mean ± Std. Dev.	P Value	Satisfaction Correlation
Sit To Stand (s)	Section 5.5.1	1.03 ± 0.53	1.8 ± 1.22	.001	-0.49
Stand To Sit (s)	Section 5.5.1	1.81 ± 0.59	2.51 ± 1.11	.002	-0.45
Total Time (s)	Section 5.5.1	12.24 ± 3.5	16.96 ± 6.65	.001	-0.60
Turning At Goal (s)	Section 5.5.1	0.59 ± 0.24	0.71 ± 0.23	.114	-0.26
Walking To Chair (s)	Section 5.5.1	4.58 ± 1.55	6.14 ± 2.82	.008	-0.50
Walking To Goal (s)	Section 5.5.1	3.49 ± 1.04	4.60 ± 1.57	.003	-0.58
Non-op End Flexion (deg)	Section 5.5.2	84.15 ± 12.01	75.43 ± 13.15	.027	0.12
Non-op Accumulated Flexion (deg)	Section 5.5.4	774.21 ± 131.94	836.59 ± 177.47	.164	-0.33
Non-op Accumulated Shin (deg)	Section 5.5.3	1167.64 ± 116.91	1252.20 ± 167.56	.038	-0.39
Non-op Accumulated Rotation (deg)	Section 5.5.4	443.23 ± 108.95	562.86 ± 203.45	.004	-0.40
Non-op Accumulated Thigh (deg)	Section 5.5.3	1045.77 ± 106.52	1100.95 ± 132.52	.121	-0.32
Op Accumulated Flexion (deg)	Section 5.5.4	742.8 ± 118.73	810.94 ± 142.65	.084	-0.36
Op Accumulated Rotation (deg)	Section 5.5.4	419.87 ± 134.97	505.06 ± 150.35	.054	-0.25
Op Accumulated Varus (deg)	Section 5.5.4	243.52 ± 61.15	309.10 ± 120.36	.006	-0.23
Op Accumulated Shin (deg)	Section 5.5.3	1158.68 ± 119.01	1241.48 ± 168.82	.045	-0.39
Op Accumulated Thigh (deg)	Section 5.5.3	1028.61 ± 100.47	1083.63 ± 117.11	.096	-0.33
Non-op Step Accumulated Flexion (deg)	Section 5.5.5	92.05 ± 18.04	83.01 ± 18.93	.120	0.24
Non-op Step Accumulated Shin (deg)	Section 5.5.5	132.08 ± 21.20	113.12 ± 20.03	.006	0.45
Non-op Step Accumulated Thigh (deg)	Section 5.5.5	93.82 ± 20.27	77.06 ± 12.23	.007	0.43
Non-op Step Max Flexion Velocity (deg)	Section 5.5.5	297.32 ± 76.69	247.62 ± 61.87	.039	0.25
Op Step Accumulated Flexion (deg)	Section 5.5.5	87.57 ± 14.88	78.20 ± 16.08	.054	0.30
Op Step Accumulated Shin (deg)	Section 5.5.5	131.09 ± 20.88	112.73 ± 23.50	.008	0.45
Op Step Accumulated Thigh (deg)	Section 5.5.5	92.51 ± 18.83	76.20 ± 17.14	.007	0.46
Op Step Max Flexion Velocity (deg)	Section 5.5.5	286.31 ± 68.38	238.12 ± 63.58	.027	0.26

Performing PCA indicated 20 principal axes were needed to explain 95% of the variance in patient function.

5.6.4   Discussion

Test segmentation and the thresholds used in the current work were sufficient to segment all patient tests and were verified using visual recreation of the tests using the leg mesh avatar (Figure 5.2). Further extraction of functional parameters that can be used to evaluate patients following total knee replacement revealed many parameters that were different between satisfied and dissatisfied groups. From Table 5.9 it can be seen that all self-reported measures were also significantly different between satisfaction groups which supports previous work indicating PROMs are likely to suffer from ceiling and floor effects, where if a patient is dissatisfied they are likely to approach the remainder of the measures with more negative bias.

Self-reported functional evaluation may not be capable of distinguishing any specific deficiencies that may be leading to functional dissatisfaction, the extracted measures show promise to aid in diagnosing joint problems.

Many metrics indicating the efficiency of test execution have been shown to correlate with satisfaction. From Table 5.10 it can be seen that the traditional total time metric is most correlated which represents a general measure of test function. Because of further metrics extraction, we can see that accumulated metrics from the Section 5.5.3 and Section 5.5.4 categories show weak to fair negative correlations in operative and non-operative legs. Lower overall amounts of motion in the shin and thigh sensors could indicate more efficient test execution with less unnecessary effort expended. When examining step metrics described in Section 5.5.5, positive weak to fair correlations are observed in accumulated motion metrics for both operative and non-operative legs. From other literature, an increased range of joint motion correlates with satisfaction post-surgery and this is reflected with these positive correlations for separated step activities. Maximum flexion metrics during walking segments did not correlate to satisfaction in the current work and there is little evidence to show that large maximum flexion ranges during walking indicate positive function. It is likely that patients have a larger maximum flexion capability than they traverse during normal walking.

The PCA revealed that 20 principle axes were required to describe 95% of the variance in metrics among the patient population. Due to the trend of lower test times and lower overall motion metrics indicating more satisfied patients, it may be expected that metrics are linearly dependent. Since the variance in patient performance could not be explained with a small number of axes, this is not the case. This indicates instead that there are many functional differences in test execution that were measured by many parameters and dimensionality reduction would not be overly successful. It should also be noted that single metrics should not be expected to be highly correlated to patient outcomes. If there were only a few functional parameters that could be observed by all patients that would indicated positive or negative outcomes, it would be expected that these could be found with simple observation and further sensor instrumentation would not be required. It is instead expected that a combination of metrics in conjunction with patient demographics would be most useful for outcome evaluation. For example, an elderly or obese patient will likely expect different functional performance to be satisfied with

their surgery than a young person who may have a much more active lifestyle. Future work should examine sensor-measured functional performance in larger populations to investigate acceptable functional states for surgical recovery in stratified groups [118].

It is also expected that instantaneous function may correlate roughly to satisfaction but a very relevant portion of a patient's satisfaction may come from the amount of improvement they have shown. It would be expected that a patient with extremely poor function that has improved to lower than average would still be satisfied and patients that experienced little improvement but show good function may be dissatisfied at the current timepoint of examination. The self-reported satisfaction score is inherently tied to past patient experiences and will suffer this bias. Future work should examine changes in the derived metrics across patient recovery, or changes from before surgery to more appropriately indicate patient improvement and incorporate the amount of recovery into parameters used to evaluate surgical success or indicate undesirable outcomes along the patient's recovery path. This idea is backed by other work indicating patient reported acceptable function state varied across patient baseline scores much more than across age, disease duration, or sex [119]. Future extensions of this analysis could include metric sensitivity testing by instrumenting the same patients for multiple tests to identify more granular intra-subject variation of test completion. This analysis could help reveal parameters that may be clinically important and could also identify some that tend to vary greatly even between the same subject in sequential tests.

As an early stage of processing, the start and end point of tests are automatically detected. In the current work this has been leveraged to remove observer bias and remove any cognitive delays between when the tester indicates "go" and when the subjects begin their test. This could also offer the ability for subjects to self-test, without an external observer for objective self-evaluation and assessment. Given that subjects can easily attach and remove the hook and latch straps themselves from a seated position, it could be expected that they could complete tests remotely, possibly with less need for in-clinic visits to ensure function is as expected and still maintain low-bias measurements.

Although recreated animations were effective for evaluating test segmentation success, there are limitations with the assumptions made about the start position. Due to the slight errors in the shape of subjects' thighs (some morbidly obese), this assumption that their thigh

sensors would be approximately parallel to the floor did not always hold true. The effect of this is a visual tilt in the remainder of the test as if the subject was walking on a slight incline. Since the animation was for visual confirmation only, it was not a result that negatively impacted evaluation.

During segmentation of the TUG, tests were first trimmed to a motionless state to ensure participants are ready to begin the test and have not started recording before becoming seated in the starting chair. The true test start is determined as the first signs of motion following this to capture starting strategies such as flexing a single leg to pre-load standing momentum however, the current analysis has focused on activity starting at the stand. Future work should examine the effect of increasing flexion before standing in the operative and non-operative leg to look for patterns in more stiff knee joints to determine if this could be an important parameter indicating poorer function.

5.7 Contribution

The motivation for this work includes real time test segmentation and data extraction for use in the clinic although the current analysis has been done during post-processing of the subject data. Future improvements on this work will include deploying these algorithms to the portable system so that the information can be available to clinicians directly following patient testing so they can provide appropriate suggestions in real-time. Contributions of the current chapter consist of novel methods to extract functional joint parameters from easy-to-complete functional tests. The metrics developed have shown early indications that they can separate patient function but further utility of the metrics will be explored in upcoming chapters. The majority of this content is currently under review for publication in *Medical Engineering and Physics*.

Chapter 6

Predicting Functional Recovery

6.1 Introduction

Following the derivation of many granular metrics to describe patient functional capability measured during the TUG test, further analysis was required to determine the clinical relevance of these additional measures. The joint measurements and test metrics derived in Chapter 5 using the developed wearable sensor system in Chapter 4 offer abundant data with numerous parameters and without further interpretation, their utility is limited. Due to the complexity these measures, patterns of desirable functional traits across patient performance and recovery are not easily identifiable but machine learning offers the ability to identify complex multivariate patterns in data and can group similar patient populations together without pre-defining group labels or membership criteria. Previous work in Chapter 5 has looked at correlations between PROMs and patient outcomes at a single post-surgery follow-up timepoint. By instrumenting patients at several time points pre/post surgery there is an abundance of data available which previously were not practical to acquire. It is expected that the additional TUG information paired with traditional PROMs will permit more accurate prediction models to be trained. Ultimately it is hoped that outcome models can be used clinically to group patients into various risk categories and will allow post-surgical resources to be planned more accordingly to ensure the best care is provided to patients.

The purpose of the work demonstrated in this chapter was to refine and filter sensor-derived measures by continuing to instrument TKA patients before and during their short-term recov-

ery period and apply machine learning to help parse recovery patterns. Firstly, unsupervised machine learning was used to organically separate patients into different functional clusters without defining any group membership criteria. These patient groups were analyzed by comparing the functional status and recovery between groups to determine traits common among patients with desirable recovery paths. Patients were instrumented and examined at several longitudinal timepoints through their recovery. All clinical instrumentation was completed by Harley Williams and Jordan Broberg as a method of functional evaluation in their studies concerning patient outcomes following knee replacement.

Analysis considered the first three months of recovery, known as the early recovery period. This time is clinically relevant for correcting undesirable recovery traits through early surgeon intervention and physiotherapy, but also for health economics and information provided in this time can have significant impact on patient health throughout the remainder of recovery. Additionally, preoperative test results will be primarily examined since the prediction of patient recovery based on functional performance is most purposeful for appropriately allocating knee replacement resources to deal with potentially troubling cases. PROMs will be collected in addition as the standard of care tool for evaluating patients and their recovery success.

Following linking preoperative function as measured by derived metrics to future patient recovery, the remainder of the chapter details the implementation of machine learning models to accurately predict if a patient will experience a clinically relevant improvement through functional testing. Knowing if a patient will or will not exhibit substantial improvements before undergoing surgery can help set realistic preoperative expectations, which are a major contributing factor to low satisfaction post-surgery [4].

6.2 Unsupervised Preoperative Clustering

6.2.1 Patient Population

Patients undergoing unilateral TKR as a treatment for OA were pooled from two separate, ongoing studies that included instrumentation and measurement procedures using the developed wearable sensor system as secondary objective method of functional evaluation. Both

studies were approved by the institutional ethics review board and informed consent was collected from all patients prior to participation at their pre-admission clinical appointment (Appendix A). All surgeries were performed at LHSC Rorabeck Bourne Joint Replacement Clinic at University Hospital. Before being approached for enrolment, all patients were pre-screened using an orthopaedic database to obtain demographics and ensure they did not have inflammatory arthritis, neuromuscular diseases, alcoholism, or language and/or cognitive barriers. Study A excluded patients older than 75 years of age while study B excluded patients with a BMI of over 40 kg/m^2 or a varus/valgus knee alignment offset greater than 15 deg. By pooling patients from studies A and B, a study population was obtained ranging in age, BMI, and function.

6.2.2 Instrumentation Procedure

At preoperative clinical appointments, all patients completed three trials of the TUG test while instrumented with the developed wearable sensor system to measure joint performance of both operative and non-operative knees. To efficiently integrate testing into the standard of care, tests were recorded in pre-admission or orthopaedic clinic hallways using a standardized waiting room chair, a 3m tape mark on the floor, our minimal sensor system, with no additional equipment. In addition, facility-standard PROMs were collected which included: The Short Form 12 (SF-12), Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC), Knee Society Score (KSS), and University of California Los Angeles (UCLA) Activity Score.

Patients followed the same protocol at their two and six week, three and six month, and one year clinical follow-up appointments. Data were transferred wirelessly (Bluetooth 4.0) from the inertial sensors to an Apple iPod Touch 5G (CA, USA) as raw orientation measurements at a rate of 25 Hz, which has been shown to be of sufficient rate to capture lower extremities [104]. The orientations of the upper and lower leg segments and their respective differences were used to compute 55 joint-specific and spatiotemporal metrics to objectively measure the TUG test as detailed in Chapter 5 and all tests were autonomously segmented into TUG sub-activities using patient spatial information and the known test structure. Tester bias was removed from recordings by automatically detecting the test start and end times during post-processing using motion thresholds.

6.2.3 Methods

The homogeneity of the combined study population from studies A and B was confirmed by performing multiple Mann-Whitney tests (Prism 8, GraphPad) on preoperative PROMs, age, and BMI, to independently identify baseline differences in each measure between included study groups A and B and a Fisher's exact test was used to detect a difference in sex. These tests were used to ensure both study populations had comparable PROMs and differed only in their study exclusion criteria.

Extracted functional and spatiotemporal metrics from patient preoperative tests were combined into feature rows with one row indicating a preoperative patient sample. These samples were stacked to create a dataset of preoperative metrics from patients with varying functional ability. Each feature column was standardized to have zero mean and unit standard deviation, so all features varied on a similar scale.

Since it was not initially known which metrics were indicative of positive or negative functional traits, unsupervised machine learning was chosen to separate groups of patients. All preoperative samples were fed into an unsupervised K-means clustering algorithm which separated patients into two groups. Binary groups were not decided initially but this ideal number of groups was validated by observing a minimal Calinski-Harabasz index (with k=2, tested on a range of 2-6).

Cluster Analysis

A one-way multivariate ANOVA was used to find PROM differences between the two separated clusters and Anderson-Darlington tests were used to successfully confirm the normality of all preoperative PROM distributions (p <0.05). One-way ANOVA was also used to find significant differences in the derived spatiotemporal and functional metrics between groups. Additionally, Cohen's D was computed for each feature to determine which of the derived metrics had the largest effect size and contributed the most to differentiating the two clusters.

The preoperative functional clusters were tracked forwards to their twelve-week appointments and a mixed-effects ANOVA with repeated measures test was performed on TUG total completion times of both groups at the preoperative, two, six, and twelve-week timepoints.

Since the same subjects were analyze at multiple timepoints, each timepoint was a factor of the repeated measures analysis. Derived metrics that remained significantly different between groups at both preoperative and twelve-week timepoints with the largest effect sizes were observed as the most important persistent functional differences and were also analyzed using a mixed-effects ANOVA with repeated measures to compare improvement between preoperative and twelve-week trials.

6.2.4 Results

A total of 94 patients were eligible to participate in this study but 26 were excluded because they missed/rescheduled their preoperative appointment or did not remain in the study up to their twelve-week appointment. The remaining 68 patients (M:F = 34:34) were included in the study aged 67.5 ± 9.8 years with a BMI of 33.5 ± 6.0. Clustering feature rows for all preoperative TUG trials (three per patient) partitioned patients into two groups with sizes 46 and 22 (M:F=20:26 and M:F=14:8). Seven patients had at least one TUG trial sorted into each group but were separated into the group with majority membership. Comparison of spatiotemporal metrics revealed the mean total TUG completion time (the only traditionally recorded TUG metric) of the larger group's trials was significantly faster than the smaller group (12.6s vs. 21.6s, $p<.001$) but there was a large overlap in trial times between groups (Figure 6.1).
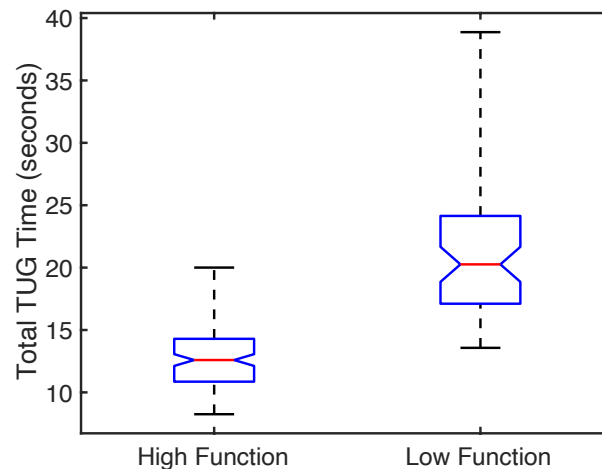
Figure 6.1: Box and whisker plot of total TUG completion time of both groups preoperatively showing a large overlap in times between groups.

Despite the large group containing slow trials and the small group containing fast trials,

these two groups were labelled "high function" and "low function" respectively due to their mean total TUG time. The high function group was younger, scored higher in UCLA, KSS Functional Activities, Office Knee Evaluation Function (surgeon reported), and Office Knee Evaluation Total (surgeon reported) questionnaires and had slightly higher (KSS) expectations (Table 6.1).

Table 6.1: Mean ± standard deviation values of patient characteristics and questionnaire outcomes of high and low function preoperative clusters. Asterisk (*) indicates surgeon reported measures and plus (+) indicates differences above MCID.

Mean ± SD	High Function	Low Function	P Value
Age (*years*)	**65.6 ± 9.1**	**71.5 ± 10.3**	**p = 0.018**
BMI (kg/m^2)	32.8 ± 5.8	34.9 ± 6.3	p = 0.180
UCLA Activity Score[+]	**4.9 ± 2.1**	**3.8 ± 1.6**	**p = 0.001**
SF-12 Mental	53.3 ± 11.0	51.7 ± 9.8	p = 0.493
SF-12 Physical	33.1 ± 8.3	29.7 ± 8.9	p = 0.079
WOMAC Pain	46.6 ± 17.3	45.5 ± 17.9	p = 0.682
WOMAC Stiffness	41.2 ± 19.0	46.2 ± 24.3	p = 0.123
WOMAC Function	49.3 ± 18.7	49.8 ± 11.0	p = 0.838
WOMAC Total	46.1 ± 15.8	47.2 ± 13.8	p = 0.620
KSS Symptoms	16.0 ± 4.1	15.3 ± 3.8	p = 0.246
KSS Satisfaction	13.7 ± 7.6	14.5 ± 7.1	p = 0.506
KSS Expectations	**14.0 ± 1.3**	**13.0 ± 2.0**	**p <0.001**
KSS Functional Activities	**37.9 ± 16.7**	**32.3 ± 12.5**	**p = 0.029**
KSS Knee Objective Indicators[*]	34.4 ± 18.5	33.8 ± 17.8	p = 0.813
Knee Evaluation Function[*]	**50.7 ± 13.6**	**41.8 ± 23.1**	**p = 0.007**
Knee Evaluation Total Knee[*]	43.0 ± 15.3	39.3 ± 16.1	p = 0.201
Knee Evaluation Total[*]	**93.7 ± 25.4**	**78.1 ± 33.0**	**p = 0.003**

Most of the novel derived measures (48/55) were also significantly different between groups (p<0.05) including all spatiotemporal metrics (Table 6.2) which is expected since the k-means algorithm separates clusters to achieve a maximum separation across all features equally. The ten derived metrics with the largest effect size can be seen in Table 6.3.

The mean total TUG time for sorted patients was not only different between groups preoperatively, but also at the six (5.4s, p=0.01) and twelve (4.2s, p=0.02) week follow-ups. Total time was not significantly different between groups at the two-week timepoint (p=0.55) where the mean patient time increased for both high and low function groups (by 9.2s and 7.6s, respectively). Mean total time of the high function group improved 0.87s (p=0.07) from preoperative

Table 6.2: Mean ± standard deviation of spatiotemporal metrics differences between preoperative patient clusters.

Mean ± SD	High Function	Low Function	P Value
Total Time	12.7 ± 2.4	21.6 ± 5.7	p<0.0001
Sit to Stand	1.1 ± 0.5	2.2 ± 1.3	p<0.0001
Walking to Goal	3.8 ± 0.8	6.6 ± 2.0	p<0.0001
Turning at Goal	0.6 ± 0.3	0.4 ± 0.5	p<0.0001
Walking to Chair	4.8 ± 1.0	8.1 ± 2.0	p<0.0001
Stand to Sit	1.8 ± 0.5	2.8 ± 1.5	p<0.0001

Table 6.3: Mean ± standard deviation of top distinguishable functional and spatiotemporal metrics between groups at pre-operation and their effect size (Cohen's D). All features are significantly different between groups (p<0.001).

Metric Description	High Function	Low Function	D
Time taken walking back to test start after turning (s)	4.8 ± 1.0	8.1 ± 2.0	1.604
Mean additive operative lower leg motion during steps	124.5 ± 14.4	92.1± 12.0	1.601
Total test time (s)	12.6 ± 2.4	21.6 ± 5.7	1.591
Mean additive non-operative lower leg motion during steps	128.7 ± 15.4	96.0 ± 11.2	1.571
Time taken walking from initial stand to begin of turn (s)	3.8 ± 0.8	6.6 ± 2.0	1.510
Mean additive operative upper leg motion during steps	90.0 ± 14.7	63.4 ± 9.8	1.470
Mean additive non-operative upper leg motion during steps	91.1 ± 14.9	65.1 ± 7.6	1.468
Mean non-operative step peak flexion velocity (deg/s^2)	289.9 ± 57.2	196.2 ± 22.0	1.428
Mean non-operative step peak extension velocity (deg/s^2)	282.3 ± 59.5	197.8 ± 37.6	1.271
Mean non-operative step peak flexion acceleration (deg/s^2)	5412.1 ± 1692.7	3129.9 ± 895.6	1.247

to twelve weeks (Table 6.4) while the low function group improved by 4.94s (p=0.005). Of the patients in the high function group, 26% (12) had meaningfully improved function (completion time decreased by >2.27 s from pre-operation to twelve weeks post-operation), 63% (29) maintained function, and 11% (5) had worsened function (completion time increased by >2.27 s). This threshold of 2.27s has been found to represent a clinically meaningful change in overal TUG test completion time in patients recovering from knee replacement [116]. Note

that since the instrumentation used in this work records measurements at only 25 Hz, this level of precision will not be reached during automatic segmentation of tests. For comparison of this work to other work in the field, this threshold of 2.27s will be used in the remaining work under the assumption that the current system can reach approximate this target. Of the patients in the low function group, 64% (14) had improved function, 27% (6) maintained function, and 9% (2) had worsened function. A comparison of individual TUG segment times for all trials of both groups at each time point can be seen in Figure 6.2.

Table 6.4: Mean functional group TUG total completion time changes and confidence interval between preoperative performance and each recovery point (negative values indicate a total time improvement).

Timepoint Comparison	Maintainers (High Function)			Responders (Low Function)		
	Mean	95% CI	P Value	Mean	95% CI	P Value
Preoperative to 2 Weeks	+9.2 s	+5.1 to +13.3	<.001	+7.6 s	+6.1 to +21.4	.400
Preoperative to 6 Weeks	+0.4 s	-1.0 to +1.8	.472	-2.9 s	-7.9 to +2.1	.373
Preoperative to 3 Months	**-0.9 s**	**-1.8 to +0.1**	**.076**	**-4.9 s**	**-8.5 to -1.3**	**.005**
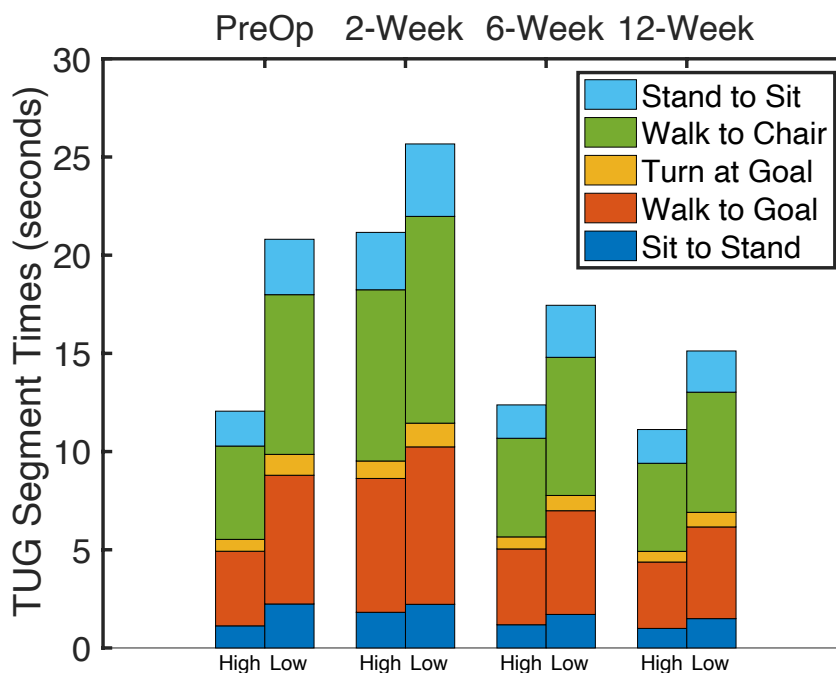
Figure 6.2: Mean TUG segment times for each functional group over their early recovery period.

The most distinguishable features between the two groups at twelve weeks and their effect sizes can be seen in Table 6.5. The top three most distinguishable features at the twelve-week

timepoint (top of Table 6.5) have been plotted alongside the same dimensions of the two cluster centroids found using all 55 metrics to visualize their strong influence on the preoperative group separation (Figure 6.3). Improvement of key metrics for both groups can be seen in Table 6.6.

Table 6.5: Mean ± standard deviation of top distinguishable functional and spatiotemporal metrics between groups at twelve weeks post-operation and their effect size (Cohen's D). All features were significantly different between groups (p<0.001). Bold indicates features that were also distinguished in the preoperative timepoint analysis.

Metric Description	High Function	Low Function	D
Mean additive operative upper leg motion during steps	**96.1 ± 15.0**	**76.8 ± 18.9**	**1.043**
Mean additive non-operative lower leg motion during steps	**135.0 ± 15.4**	**114.8 ± 22.2**	**1.004**
Mean additive operative lower leg motion during steps	**131.2 ± 16.3**	**109.5 ± 27.1**	**0.961**
Time taken walking from initial stand to begin of turn (s)	**3.4 ± 0.7**	**4.7 ± 2.0**	**0.933**
Mean additive non-operative upper leg motion during steps	**97.5 ± 15.1**	**81.6 ± 16.8**	**0.919**
Total test time (s)	**11.8 ± 2.7**	**16.0 ± 6.5**	**0.893**
Time taken walking back to test start after turning (s)	**4.5 ± 1.2**	**6.1 ± 2.5**	**0.868**
Time taken to stand from the seated start position (s)	1.7 ± 0.7	2.1 ± 0.9	0.800
Mean additive operative flexion during steps (deg)	87.0 16.1	73.9 ± 15.2	0.774
Mean operative flexion range during steps (deg)	42.4 ± 7.9	36.5 ± 7.5	0.714

Table 6.6: Operative and non-operative improvement of top persisting motion metrics between preoperative and twelve-week follow-ups for each group.

Metric Description	High Function			Low Function		
	Mean	95% CI	P Value	Mean	95% CI	P Value
Mean additive operative upper leg motion during steps	6.1	10.2 to 2.0	p=0.002	13.3	19.5 to 7.1	<0.0001
Mean additive non-operative lower leg motion during steps	6.3	10.7 to 1.9	p=0.003	18.8	25.3 to 12.4	<0.0001
Mean additive operative lower leg motion during steps	6.6	11.4 to 1.8	p=0.004	17.4	24.6 to 10.2	<0.0001
Mean additive non-operative upper leg motion during steps	6.4	10.3 to 2.4	p=0.001	16.5	22.3 to 10.7	<0.0001
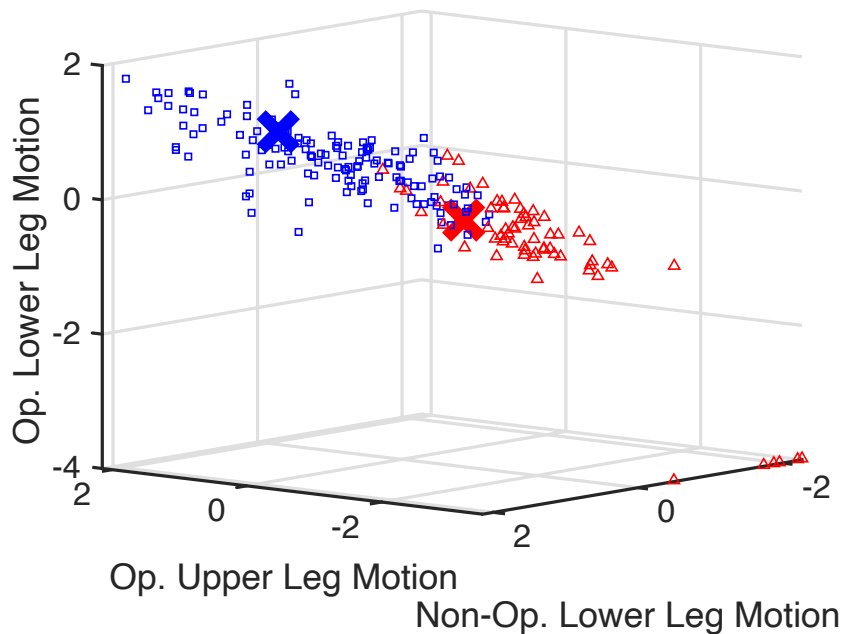
Figure 6.3: Z-scores of the top three most distinguished metrics persisting to the twelve-week follow-up appointments for each trial and their influence on initial preoperative clustering. Solid crosses (x) represent group centroids found using all 55 derived metrics. Red triangles and blue squares indicate high and low function group trials respectively.

6.2.5 Discussion

The unsupervised clustering performed in this study successfully separated wearable sensor instrumented performance tests based on derived functional metrics into clinically relevant groups. Literature has previously linked preoperative functional performance to postoperative functional improvement, however, this study has highlighted functional parameters that differentiate between patients that are more likely to show functional improvement [83, 120].

The objective measures extracted from the short and easy-to-implement TUG test have been able to distinguish function preoperatively with many significant differences between groups whereas there are few differences in PROMs and overlapping TUG completion times. It can be seen in Table 6.1 that despite several PROMs being significantly different between groups, only the UCLA Activity Score varied more than its minimal clinically important difference (MCID) of 0.92 [121, 122]. This presents evidence that there are functional performance differences that cannot be distinguished using subjective self-reported measures when comparing patients with different expectations.

Tracking the two functionally separated groups through to three months revealed that one

group was much more likely to improve relative to the other. Test results for new patients can be compared to the two groups clustered in this study to determine the most similar expected recovery path, and this information can be used to provide realistic recovery expectations for patients. The analysis performed has relied on total TUG time as a relatable overall performance metric and tool for labelling groups but it should be highlighted that other derived metrics had larger effect sizes between groups and were determined to be more relevant parameters influencing the resulting group memberships (Table 6.3, Table 6.5). Identified features distinguishing the functional groups at twelve weeks included both operative and non-operative leg metrics which suggests test completion and movement compensation strategies were captured. The high improvement group had larger improvements in all relevant additive motion metrics (Table 6.6) but these values never exceeded those of the faster group. Increased ROM has been previously linked to better outcomes but the results of this study indicate that the total additive amount of motion expended during activities by both op- and non-operative legs is also functionally important [14, 123]. Both groups improved these metrics significantly during recovery which supports that the metrics are influencing shorter TUG times, which did decrease for both groups but not by a meaningful difference for the high function group.

Although it may be thought that higher functioning preoperative patients have less possible function to regain and will likely show less functional improvement, there is still a benefit of including the instrumented TUG test at preoperative visits since this functional differentiation was possible with the derived metrics and the current work has shown that PROMs alone cannot reliably report function to this granularity. Additionally, it is important to note in the preoperative groups that seven patients in the "low" function group had mean total TUG times faster than the worst "high" function time. Similarly, there were seven patients in the "high" function group with times slower than the best "low" function group time, indicating that total TUG time alone would not be enough to sort the groups in this way. Of the seven patients in the "low" function group with favourable times that have been labelled likely to improve, five of them have improved their total time above the TUG MCID of 2.27s [116]. Of the seven patients in the "high" function group with less favourable times labelled not likely to improve, only three of them have improved above the same threshold. These patients that were sorted into the low function group who had more favourable times have still shown improvement

despite a smaller possible improvement range while patients sorted into the high function group who had less favourable times have not improved despite a larger possible improvement range. These special cases suggest the cluster separation better indicates likelihood of improvement than overall functional level.

Some limitations were noted with the study performed. During data collection, seven patients repeated one of their trials sufficiently different from the others that the samples were split between both groups. Although a single patient cannot belong to multiple functional groups, the separation of tests may be a valid result since repetitions of the same test can be slightly different due to fatigue, test familiarity, confusion, or perhaps a stumble or stiffness. During clustering it was decided to keep each trial as a separate sample to best separate functional groups but when future data are compared to find the most relatable path, it may be more practical to take an average across multiple trials for more generalized predictability.

The follow-up time analyzed was limited to the early recovery period of twelve weeks. Despite recovery following TKA usually lasting one to two years, the authors believe valuable information can be obtained during the early recovery phase, and this time period can be important for health economics [124]. As alternative joint replacement payment models such as Medicare's Bundled Payments for Care Improvement Program (BPCIP) are introduced, early outcome prediction becomes valuable for allocating care costs. Under model 2 of the BPCIP, hospitals will be reimbursed for costs saved in the first 90-days following surgery and patients with early improvement will likely require less frequent early care [125]. Fast functional improvement and early ambulation reduces hospital length of stay (LOS) which also reduces the likelihood of costly readmissions due to infection [126–128]. A longer follow-up will be necessary to determine how function changes in each group until patients are fully healed.

The current work has shown that preoperative functional assessment can benefit from the use of wearable sensor instrumentation and machine learning techniques can identify multivariate patterns that would be otherwise difficult to see by an observer. Groups of patients following similar short-term recovery paths have been identified and future test data can be compared to similar path prediction to influence better patient expectations. There was little evidence that the PROMs collected in this study related to the results found using the derived functional and spatiotemporal metrics. Obtaining PROMs proved much more time consuming

for patients and the process was more cumbersome when it became time to store and digitize the measures, further motivating the use of an automated sensor system that can record performance tests in only a few minutes and provide instantaneous analysis.

6.3 Functional Recovery Classification

The work presented in Section 6.2 showed that sensor-measured performance metrics could distinguish function in preoperative knee patients where self-reported measures could not. The current work examines the use of supervised machine learning to use sensor-derived features to train and test a classifier to predict functional recovery during the early recovery period following surgery. A successful method of predicting the expected recovery of patients undergoing total knee replacement for knee osteoarthritis is beneficial for health economics, detection of functional impairment, and appropriately setting patient expectations before surgery, lowering the risk of patient dissatisfaction. Although the functionally clustered groups were roughly indicative of early recovery likelihood, it is expected that some parameters are of more importance while others are unrelated to the amount of functional recovery to be expected. An aggregate of these features can be refined to better predict functional improvement using supervised machine learning, a branch of techniques that can simply be considered as the optimization of feature combinations and scaling factors to best fit a set of labelled samples. Models developed are then tested on a withheld, unseen dataset to obtain realistic estimates of performance on new samples.

The primary outcome of this work was to refine and validate a recovery prediction strategy using preoperative functional performance to predict if a patient is likely to improve in the early recovery period or maintain their objective preoperative function, giving insight into their expected recovery trajectory. Other outcomes included determining the effect of including PROMs, age, sex, and BMI demographic information on short-term recovery predictions and if recovery trends in responders and maintainers continued to their one year follow-up, giving an estimate of long-term functional improvement.

6.3.1 Methods

Unsupervised clustering was initially chosen to separate patients organically without defining group membership criteria because it was not initially known how each metric contributed to functional distinctions. All patients available at the time of performing experiments were used and cluster analysis revealed many patients in one cluster improved their total TUG times in the early recovery period of three months (above a minimal detectable change of 2.27s) while many in the other cluster maintained their preoperative ability. Since this previous work and up to the current work, all initial study patients have completed up to their one year clinical follow-up appointment and the last of recruited study patients have completed their three month appointments. The more recently recruited patients completing their three month appointment after initial clustering (n=14) have been withheld as a test set to verify developed prediction models. Preoperative PROMs of test patients were compared to the initial clustered training set using independent Mann-Whitney tests to verify the test set had comparable baseline measures. True functional responder or maintainer labels were assigned to all patients using total TUG time as a generalized measure of function and patients with >2.27s of improvement in time were labelled as true positive responders.

To examine the recovery trend for true labelled responder and maintainer patients and determine if early recovery indicated long-term recovery, Prism 8 (Graphpad, CA, USA) was used to compute a mixed-effects model with the Geisser-Greenhouse correction and Sidak's multiple comparison tests to compare mean patient TUG times across each time point and, once again, to compare the row means of each group independently across time.

Model Development

Using the labelled patient data split into training and testing sets, three types of supervised classifiers were trained and evaluated for comparison: support vector machine (SVM), naïve Bayes (NB), and random forest (RF). Each model examined was trained and evaluated three times, first with no feature selection and two additional repetitions with a different selection scheme. The first scheme was formed by comparing the Cohen's D effect size of all features between the true labelled responders and maintainers from the training set. All features were

ranked and those with an absolute value of >0.6 were chosen for inclusion. Secondly, the genetic algorithm used by Babatunde et al. was used as a method for feature selection [129].

The training and testing of all models was repeated three additional times while expanding the feature set for patient tests to first include only sensor-derived functional metrics, then also combining age, sex, and BMI demographics, and finally also including collected PROM scores. A summary of schemes evaluated can be seen in Figure 6.4. For all models, prediction scores were extracted during evaluation to obtain a measure of classification confidence and additionally, test set classification accuracy, sensitivity, and specificity as well as the mean incorrect prediction score (where a lower value indicates less confidence), and area under the curve (AUC) of the corresponding receiver-operator curve (ROC) measures were used to measure classifier performance. AUC provides a beneficial overall measure of accuracy, especially when classes are imbalanced [130]. Although patients recently recruited into the study after the initial training set represent a valid test set, 10-fold cross validation was also performed to obtain evidence of proper model generalization on a variety of unseen patients and ensure the original test set did not represent optimistic performance. This additional validation consisted of ten iterations of randomly withholding a set of test patients from the total study population to be withheld for evaluation and using the remainder to train model parameters.

Cummulative feature set	Feature selection methods	Classifier models
1. Sensor-derived functional metrics 2. Patient demographics 3. PROMs	1. None 2. Effect sizes 3. Genetic algorithm	1. Support vector machine (SVM) 2. Naïve Bayes (NB) 3. Random forest (RF)
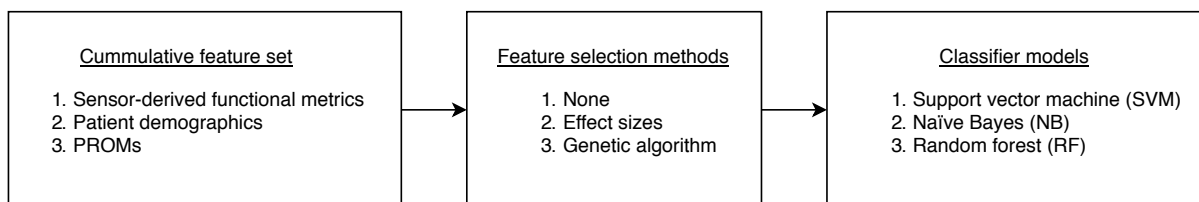
Figure 6.4: Chart describing the variables examined during training and testing of the recovery prediction classifier.

6.3.2 Results

Combining all patients from studies A and B (Figure 6.5), 119 eligible patients were recruited to participate and provided informed consent but 7 cancelled their surgeries, 5 withdrew their consent, and the surgeon deviated intraoperatively from the planned intervention for 12 patients. Of the remaining 95 patients, 82 completed both their preoperative and three month

tests.

Study A

261 assessed for eligibility

173 ineligible

- (67) Age > 75 years
- (16) Inflammatory arthritis
- (4) Alcoholism
- (6) Language barrier
- (3) Cognitive impairment
- (59) Other study recruitment
- (16) Cancelled clinical appointment
- (1) Missed contact due to scheduling
- (15) Previous joint surgery
- (3) Not able to safely ambulate

71 approached at pre-admission

- (23) Declined participation

48 recruited to the study

- (2) Surgery cancelled
- (7) Surgeon deviated intraoperatively from the planned intervention

Study B

523 assessed for eligibility

330 ineligible

- (85) BMI > 40 kg/m^2
- (66) Inflammatory arthritis
- (7) Alcoholism
- (15) Language barrier
- (5) Cognitive or neuromuscular disorder
- (75) Other study recruitment
- (36) Allergy or incompatibility with planned intervention
- (17) Cancelled clinical appointment
- (15) Missed contact due to scheduling
- (3) Not able to safely ambulate

193 approached at pre-admission

- (122) Declined participation

71 recruited to the study

- (5) Withdrew consent
- (5) Surgery cancelled
- (5) Surgeon deviated intraoperatively from the planned intervention

Pre-op (93)

- 2 missed to malfunctioning sensor

Two-week (80)

- 6 missed or rescheduled appointment
- 9 declined due to excessive pain

Six-week (87)

- 8 missed or rescheduled appointment

Three-month (85)

- 10 missed or rescheduled appointment

Six-month (82)

- 13 missed or rescheduled appointment

One-year (62)

- 15 missed or rescheduled appointment
- 18 excluded due to contralateral knee surgery

Figure 6.5: CONSORT study flow diagram including patients from the previous study with their inclusion criteria specified.

Patients from study B were older (p=.001) and had more preoperative pain (p=.023) but there were no baseline self-reported functional differences (Table 6.7). All patients who attended their scheduled clinical follow-ups were highly compliant to perform instrumented tests while they would otherwise be waiting to see their clinician with the exception of the two-week

time point where nine patients refused due to post-surgical joint pain. The 68 patients (F:M = 34:34, age = 67.5 ± 9.8 years, BMI = 33.5 ± 6.0) from our previously published work represent the original clustering and classifier training set. The additional 14 patients (F:M = 9:5, age = 68.4 ± 10.1 years, BMI = 33.5 ± 5.8) that had completed up to their three month follow-up appointment after the previous work have been withheld as a testing set. The results of Mann-Whitney and Fisher's exact tests showed no statistically significant differences in any PROMs between training and testing groups, indicating similar self-reported baseline characteristics (Table 6.8). Several patients (n=18) underwent an additional joint replacement between their six-month and one-year appointments and therefore had their one-year result excluded from the longitudinal repeated-measures analysis.

Table 6.7: Mean ± standard deviation of demographics and preoperative PROMS between both combined study groups.

Mean ± SD	Study A	Study B	P Value
Sex ($F : M$)	19:14	14:20	.225
Age (*years*)	63.2 ± 7.2	70.9 ± 10.8	.001
BMI (kg/m^2)	34.4 ± 5.9	32.2 ± 5.3	.102
UCLA Activity Score	4.7 ± 2.3	4.5 ± 1.9	.699
SF-12 Mental	52.0 ± 11.3	52.8 ± 11.3	.759
SF-12 Physical	33.0 ± 8.4	32.9 ± 8.4	.948
WOMAC Pain	40.76 ± 17.7	50.4 ± 16.4	.023
WOMAC Stiffness	41.7 ± 23.7	41.3 ± 17.5	.947
WOMAC Function	49.1 ± 19.0	49.7 ± 15.9	.883
WOMAC Total	44.0 ± 17.2	47.7 ± 13.1	.322
KSS Symptoms	15.8 ± 4.6	16.1 ± 3.5	.748
KSS Satisfaction	12.7 ± 7.7	14.6 ± 7.1	.292
KSS Expectations	13.9 ± 1.6	13.5 ± 1.5	.277
KSS Functional Activities	35.7 ± 18.5	40.0 ± 12.9	.747
KSS Knee Objective Indicators	30.5 ± 17.4	36.6 ± 19.9	.192

Clustering the 68 initial patients from the previous work resulted in two functionally distinguished groups; one predicted maintainer group (F:M = 26:20), and one predicted responder group (F:M = 8:14). Predicted responders had a significantly higher mean TUG time of 21.6s compared to the predicted maintainers time of 12.6s (p<.001) and an additional 47/55 derived functional metrics were also significantly different between groups (p<.05) [88]. Therefore, "responders" exhibited a greater functional deficit and had the potential for greater func-

Table 6.8: Mean ± standard deviation values of patient characteristics and questionnaire outcomes of initial clustered training patients and the newly included testing set.

Mean ± SD	Training Set	Testing Set	P Value
Sex ($F : M$)	34:34	9:5	.389
Age (*years*)	67.5 ± 9.8	68. 4 ± 10.1	.772
BMI (kg/m^2)	33.5 ± 6.0	33.50 ± 5.8	.992
UCLA Activity Score	4.5 ± 2.1	4.5 ± 2.0	.911
SF-12 Mental	53.0 ± 10.6	53.0 ± 7.8	.982
SF-12 Physical	32.6 ± 8.4	36.2 ± 10.4	.271
WOMAC Pain	46.5 ± 17.6	55.0 ± 15.7	.139
WOMAC Stiffness	42.5 ± 20.6	51.1 ± 23.4	.211
WOMAC Function	49.5 ± 16.5	53.4 ± 16.5	.495
WOMAC Total	46.6 ± 15.1	53.4 ± 16.8	.190
KSS Symptoms	15.8 ± 4.0	16.4 ± 4.3	.690
KSS Satisfaction	14.1 ± 7.3	18.0 ± 7.1	.115
KSS Expectations	13.6 ± 1.6	14.1 ± 1.2	.398
KSS Functional Activities	36.5 ± 15.3	45.9 ± 18.0	.112
KSS Knee Objective Indicators	34.4 ± 18.9	36.9 ± 13.4	.692

tional gains by undergoing TKA. After assigning true group labels to patients, true responders had a mean TUG time of 17.8s compared to the true maintainer group mean time of 13.3s (p=.002). The preoperative timepoint was the only one where the group mean times were different (p=.002). Beginning at the two week follow-up and persisting through to one-year, both groups had no difference in times (p>.856) and group mean differences were less than the TUG MCID of 2.27s (Table 6.9). Table 6.10 shows the results from Sidak's multiple comparisons tests for the originally clustered functional groups in the previous work up to three months compared to the true labelled responder and maintainer groups total TUG times. True responders show a clinically relevant improvement in time of 4.0s (p=.003) as early as six weeks post-operation. True maintainers were worse performing at six weeks than pre-operation on average (although the magnitude was less than the MCID of 2.27s) and did not show any meaningful time improvement even at one year post-operation (1.1s, p=.018). Only the maintainer group had significantly worse times between preoperation and two weeks (p=.001) with the maintainer group showing 67% more decline than the responders (9.5s vs. 5.7s).

Table 6.11 shows the performance of all responder classification schemes using only sensor-derived functional metrics as features when tested on the original set of newly recruited test

Table 6.9: Sidak's multiple comparison test comparing group means of total TUG times between true labelled maintainers and responders.

Timepoint Comparison	Main.	Resp.	Mean Diff.	95% CI	P Value
Preoperation	13.4 s	17.8 s	-4.4 s	-7.4 to -1.3	.002
2 Week	23.0 s	23.5 s	+0.4 s	-10.5 to +9.3	>.999
6 Weeks	14.5 s	13.8 s	+1.0 s	-2.0 to +3.4	.981
3 Months	12.5 s	12.0 s	+0.9 s	-1.6 to +2.6	.987
6 Months	11.9 s	12.0 s	+0.3 s	-2.1 to +2.0	>.999
1 Year	12.3 s	11.9 s	+0.5 s	-2.3 to +3.0	>.999

Table 6.10: Mean functional group TUG total completion time changes and confidence interval between true labelled responders and maintainers at each recovery point (negative values indicate a total time improvement). Results of the current work are shown alongside previous work.

Timepoint Comparison	Maintainers	95% CI	P Value	Responders	95% CI	P Value
Clustered Responders and Maintainers [88]						
Preoperative to 2 Weeks	+9.2 s	+5.1 to +13.3	<.001	+7.6 s	+6.1 to +21.4	.400
Preoperative to 6 Weeks	+0.4 s	-1.0 to +1.8	.472	-2.9 s	-7.9 to +2.1	.373
Preoperative to 3 Months	-0.9 s	-1.8 to +0.1	.076	-4.9 s	-8.5 to -1.3	.005
True Labelled Responders and Maintainers						
Preoperative to 2 Weeks	+9.5 s	+3.2 to +15.8	.001	+5.7 s	-1.9 to +13.3	.218
Preoperative to 6 Weeks	+1.1 s	-0.3 to 2.4	.199	-4.0 s	-6.9 to -1.2	.003
Preoperative to 3 Months	-0.9 s	-1.6 to -0.1	.011	-5.8 s	-8.4 to -3.1	<.001
Preoperative to 6 Months	-1.5 s	-2.5 to -0.4	.002	-5.8 s	-8.9 to -2.7	<.001
Preoperative to 1 Year	-1.1 s	-2.1 to 0.1	.018	-5.9 s	-9.4 to -2.3	<.001

patients. It was found that an RF classifier without any method of feature selection had the highest overall classification accuracy (0.93), sensitivity (1.00), specificity (0.75), and a very high AUC of 0.94. Using the genetic algorithm selection produced a model with an optimal AUC of 0.95, sensitivity of 0.80 and specificity of 1.00 but with a lower accuracy of 0.86. NB models were improved similarly by performing both methods of feature selection, increasing accuracy from 0.64 to 0.86, sensitivity from 0.60 to 1.00, and AUC from 0.83 to 0.88. Effect size selection was more effective for improving accuracy (0.79 to 0.86) and sensitivity (0.80 to 1.00) of the SVM models but a maximum AUC of 0.85 was observed using the genetic algorithm.

Table 6.11: Classifier performance on the valid test set of new patients withheld from model testing.

Feature Selection	Model	Acc.	Sens.	Spec.	AUC	Error Score
None	SVM	0.79	0.80	0.75	0.75	0.43 ± 0.47
	NB	0.64	0.60	0.75	0.83	0.68 ± 0.41
	RF	0.93	1.00	0.75	0.94	0.49 ± 0.00
Effect Size	SVM	0.86	1.00	0.50	0.78	0.59 ± 0.63
	NB	0.86	1.00	0.50	0.88	0.41 ± 0.10
	RF	0.71	0.80	0.50	0.83	0.55 ± 0.06
Genetic Algorithm	SVM	0.79	0.80	0.75	0.85	0.37 ± 0.23
	NB	0.86	1.00	0.50	0.88	0.42 ± 0.08
	RF	0.86	0.80	1.00	0.95	0.62 ± 0.00

Evaluation by performing 10-fold cross validation resulted in lower performance measures overall (Table 6.12), indicating this additional validation is likely more representative of deployment of these models on future populations. RF models showed the highest accuracy and AUC and using genetic algorithm selection slightly outperformed those without selection in accuracy (0.76 to 0.75), specificity (0.60 to 0.53), and AUC (0.82 to 0.80) but had reduced sensitivity (0.85 to 0.89). NB models using an effect size feature selection had the highest sensitivity of 0.92, but all SVM and NB variants were outperformed by RF in all other measures.

Continuing with 10-fold cross validated measures as a conservative estimate of performance, Table 6.13 demonstrates the benefit of including age, sex, and BMI measures into predictive models. RF models using no feature selection schemes showed the best accuracy (0.78), specificity (0.64), and AUC (0.82). NB models benefited most from effect size feature

Table 6.12: Classifier results from 10-fold cross validation using only sensor-derived functional metrics as features.

Feature Selection	Model	Accuracy	Sensitivity	Specificity	AUC	Error Score
	SVM	0.72 ± 0.11	0.89 ± 0.09	0.44 ± 0.19	0.67 ± 0.15	0.59 ± 0.39
None	NB	0.68 ± 0.10	0.79 ± 0.16	0.55 ± 0.23	0.67 ± 0.13	0.51 ± 0.42
	RF	0.75 ± 0.09	0.89 ± 0.13	0.53 ± 0.14	0.80 ± 0.14	0.43 ± 0.23
	SVM	0.71 ± 0.08	0.90 ± 0.10	0.40 ± 0.19	0.74 ± 0.09	0.42 ± 0.33
Effect Size	NB	0.69 ± 0.07	0.92 ± 0.15	0.33 ± 0.22	0.78 ± 0.09	0.45 ± 0.18
	RF	0.73 ± 0.11	0.83 ± 0.13	0.58 ± 0.27	0.75 ± 0.10	0.41 ± 0.23
	SVM	0.70 ± 0.08	0.83 ± 0.12	0.51 ± 0.22	0.71 ± 0.14	0.69 ± 0.58
Genetic Algorithm	NB	0.71 ± 0.11	0.88 ± 0.16	0.43 ± 0.26	0.79 ± 0.09	0.45 ± 0.29
	RF	0.76 ± 0.12	0.85 ± 0.11	0.60 ± 0.24	0.82 ± 0.11	0.45 ± 0.19

selection and had an accuracy of 0.73, a top sensitivity of 0.98, and AUC of 0.80 but lower specificity of 0.32. Expanding the feature set to include PROMs further increased maximum observed performance measures with the exception of sensitivity (Table 6.14). RF models using genetic algorithm feature selection showed a maximum accuracy (0.81) and AUC (0.86) with a high sensitivity of 0.93. Both SVM with no feature selection and NB using effect size selection had the highest sensitivity values (0.96) but both were outperformed by RF models in the remaining measures. While no feature selection included all PROMs in the feature set, effect size filtering removing all features with <0.60 effect size was responsible for filtering all PROMs except for the KSS functional activities and KSS satisfaction sub-scores with effect sizes of 0.66 and 0.65 respectively. The genetic algorithm selection included three PROMs with indicated effect sizes: SF12 Mental Component Score (0.30), WOMAC total (0.41), and UCLA Activity Score (0.34).

6.3.3 Discussion

Analyzing the long-term functional recovery in true responder and maintainer groups revealed only the responders observed meaningful TUG time improvements even as far as one year into recovery. While TUG times are a useful label of function, preoperative TUG times of the two true labelled groups were close together with substantial overlap in the times of individual patients, indicating TUG time is ineffective as the sole metric for separating patients. Instead, this classification must include more granular functional metrics. Despite responders

Table 6.13: Classifier results from 10-fold cross validation using sensor-derived functional metrics and patient demographics only as features. Bold indicates the most desirable value per column.

Feature Selection	Model	Accuracy	Sensitivity	Specificity	AUC	Error Score
	SVM	0.72 ± 0.07	0.89 ± 0.08	0.43 ± 0.18	0.70 ± 0.14	0.42 ± 0.34
None	NB	0.69 ± 0.09	0.83 ± 0.13	0.52 ± 0.26	0.72 ± 0.15	0.49 ± 0.38
	RF	**0.78 ± 0.11**	0.88 ± 0.22	**0.64 ± 0.20**	**0.82 ± 0.11**	0.43 ± 0.24
	SVM	0.71 ± 0.10	0.91 ± 0.16	0.39 ± 0.22	0.76 ± 0.11	0.50 ± 0.37
Effect Size	NB	0.73 ± 0.08	**0.98 ± 0.08**	0.32 ± 0.20	0.80 ± 0.07	**0.40 ± 0.07**
	RF	0.72 ± 0.11	0.85 ± 0.14	0.53 ± 0.22	0.74 ± 0.10	0.44 ± 0.25
	SVM	0.73 ± 0.12	0.90 ± 0.10	0.44 ± 0.25	0.74 ± 0.13	0.41 ± 0.31
Genetic Algorithm	NB	0.71 ± 0.09	0.88 ± 0.17	0.42 ± 0.26	0.77 ± 0.09	**0.40 ± 0.35**
	RF	0.74 ± 0.09	0.86 ± 0.09	0.54 ± 0.28	0.81 ± 0.09	0.41 ± 0.21

Table 6.14: Classifier results from 10-fold cross validation using sensor-derived functional metrics, patient demographics, and PROMs. Bold indicates the most desirable value per column.

Feature Selection	Model	Accuracy	Sensitivity	Specificity	AUC	Error Score
	SVM	0.77 ± 0.14	**0.96 ± 0.08**	0.44 ± 0.29	0.76 ± 0.14	0.44 ± 0.36
None	NB	0.73 ± 0.14	0.90 ± 0.13	0.45 ± 0.24	0.74 ± 0.15	0.41 ± 0.31
	RF	0.75 ± 0.14	0.85 ± 0.22	0.59 ± 0.30	0.78 ± 0.11	0.42 ± 0.25
	SVM	0.78 ± 0.14	0.93 ± 0.16	0.53 ± 0.24	0.80 ± 0.11	0.73 ± 0.45
Effect Size	NB	0.76 ± 0.09	**0.96 ± 0.08**	0.41 ± 0.18	0.81 ± 0.07	**0.32 ± 0.27**
	RF	0.77 ± 0.13	0.86 ± 0.14	**0.66 ± 0.27**	0.82 ± 0.10	0.45 ± 0.24
	SVM	0.71 ± 0.08	0.83 ± 0.10	0.53 ± 0.32	0.73 ± 0.13	0.74 ± 0.96
Genetic Algorithm	NB	0.73 ± 0.10	0.87 ± 0.17	0.45 ± 0.24	0.73 ± 0.09	0.45 ± 0.31
	RF	**0.81 ± 0.10**	0.93 ± 0.09	0.61 ± 0.24	**0.86 ± 0.09**	0.37 ± 0.20

being labelled at the three month timepoint, significant improvement was noted as early as six weeks post-operation. Neither group demonstrated functional changes above the MCID of 2.27s from three months on to one year. This indicates functional changes observed during the TUG test should be realised by this point, especially if the patient was predicted to be a functional responder.

The three month clinical follow-up remains an important milestone of recovery. Early identification of functional joint deficiencies provides an opportunity for alternative intervention before problems progress and require revision surgery. For example, a common treatment to restore joint range of motion after persisting joint stiffness following TKR is manipulation under anesthesia. This procedure involves forcing the knee to flex, breaking scar tissue, al-

lowing the joint to move more freely. It has been shown that this procedure is most effective when performed in the first three months following surgery [131]. If expected functional improvement has not been attained as predicted from preoperative classifications, and the physical exam and patient complaints are consistent with joint stiffness, this could be found sufficiently early that there is still an option to remedy with a relatively minor procedure. Secondly, the three months following surgery are important for health economics in new bundled care programs being deployed in the United States and Canada [132]. Under these models, the cost of surgery is bundled with all patient expenses for 90 days following surgery. This is in effort for hospitals and clinicians to provide more effective, efficient care to patients. Recovery predictions that can be made for this period could be used to effectively plan health resources to deal with varying cases of patients.

Using 10-fold validation, it was observed that testing only using our withheld test set (Table 6.11) presented an optimistic measure of performance. It is expected that the cross validated results represent a more generalized measure of performance when deployed to make predictions for future patients. Classification accuracy was found to be high in top performing models: 0.76 using only functional metrics, 0.78 after incorporating demographics, and 0.81 with the inclusion of PROMs. With the high degree of classification accuracy presented in the current work, the functional recovery predictions made using these preoperative TUG tests could be used to appropriately adjust patient expectations of postoperative functional abilities before they undergo surgery. Considering the adjustment of patient expectations, the sensitivity metric should be important since responders were labelled using a positive label, and incorrect positive labels could have a negative effect on expectations. It would be expected that a patient predicted to be a responder that did not see functional improvement in the TUG test would have unmet expectations, and thus be dissatisfied with their surgical outcome. A high sensitivity to detect true responders is important for surgeons as well, since a patient not attaining their expected recovery may require additional attention. Alternatively, a maintainer labelled patient that exhibits strong functional improvement may feel that they have exceeded expectations, with a potentially positive effect despite an incorrect prediction. For this reason, an optimal classifier will favour a high sensitivity value over specificity.

There was only a small drop in classifier performance when excluding patient reported out-

come measures. It is very time consuming for patients to complete these lengthy questionnaires and many times they are filled in incorrectly, rendering the computed measures inconclusive. Both methods of feature selection performed removed most PROMs, indicating that if they were to be included in a deployed model, it may be more practical to only collect the specific measures used for classification. Moving forward with future work predicting responders or maintainers with only sensor-system derived metrics could benefit best from a random forest model.

Several limitations could be noted with this study and analysis, including the selection of classifier algorithms. Three common classifiers were chosen for testing due to their previous success in orthopaedic research [133–135]. Artificial neural networks have not been tested for classification success since the number of patient samples is relatively low to learn complex non-linear relations using many parameters. It is expected that as more patients are instrumented by the system, stratified groups will emerge, producing more refined clusters beyond simply "responders" or "maintainers". An example of stratification in our groups could be patients with severe bilateral OA versus healthy contralateral joints. Several patients in this study (n=18) were excluded from one-year analysis due to receiving another joint replacement on the contralateral leg. With more patients, further investigation could involve separating patients who undergo their first joint replacement opposed to those with existing contralateral implants to see if recovery differs. This could also involve hip replacements, as the hip is also highly relevant to the mobility of the lower extremities, and hip impairments may be reflected in sensor-derived TUG measures.

6.4 Contribution

This study has demonstrated that the substantial additional data derived from preoperative sensor-derived metrics recorded while patients participate in the TUG test can be used to accurately predict if a patient is going to improve their test time in the early recovery period following TKR. Although only test time improvements are predicted, the total TUG time is used as an accepted measure of function in patients with knee OA [23,24]. Additional granular sensor-measured metrics were shown to be more effective to localize and distinguish patient

function predictive of early recovery compared to the overall test time measure traditionally used. This work has also shown that the future deployment of a recovery prediction classifier could benefit slightly from the inclusion of subjective measures but the performance gain may be negligible when considering logistics of collecting full PROMs from patients and combining them with test results before making real-time recovery predictions in the clinic. Excluding PROMs also removes subjectivity, making this the first entirely objective tool to predict patient function after TKR. Much of the unsupervised clustering work in this chapter has been published in the *Journal of Arthroplasty* and the model development and testing was published as additional work in *Medical Engineering and Physics* [136, 137].

Chapter 7

Activity Recognition

Instrumenting patients with the system detailed in Chapter 4 has shown great support for in-clinic measurement of patient function to accompany standard self-reported assessment measures. Evidence for easy-to-deploy functional instrumentation in the clinic has been present in Chapter 5 and Chapter 6. Due to many recent challenges associated with in-person clinical visits, an additional ideal application of wearable sensors would be to help remotely assess patient function when in-person visits are not possible. Although it is not expected that sensor integration will ever completely replace valuable feedback from surgeons or clinicians, it could be that functional assessment will provide patients with individualized feedback concerning their expected function. Similarly, surgeons or clinicians may be able to stay updated on their patients' health more distantly or to check if an in-person appointment will be required.

Furthermore, a limitation of in-clinic functional testing is that patients are being tested outside of their normal environments. The function that is demonstrated under observation may not be what patients are capable of during daily activity. Extending instrumentation to more long-term sessions would promote natural measurement and provide a realistic day-to-day depiction of functional ability opposed to short term capability demonstrated in the lab.

Unfortunately, during unscripted monitoring, further analysis of individual activities and comparison of activity-specific performance metrics across multiple subjects is not possible since it is not known what activities are being performed or attempted. More challenging tasks such as ascending or descending stairs should be analyzed differently than level-ground walking or less joint-stressing functions. Recording sessions can be given context by apply-

ing activity recognition or classification techniques which if successful, can then be used for subject-to-subject or subject-to-population comparisons for specific activities. Patient measurement from the previous chapters has relied on rigid structured tests so knee performance during similar activities could be compared between patients.

The ability to classify activities performed by an instrumented subject during extended unscripted monitoring sessions would give context to acquired measurements and will permit appropriate comparison of functional performance across subjects in many remote health assessment or evaluation applications. The configuration of four sensors to be used (one mounted above and below each knee) has been shown effective for deriving knee specific objective measures but the current work will explore the capability to extend this to more unscripted activities. Recall from Chapter 3 that recent literature has shown automatic feature extraction methods can often out-perform classification tasks using manually engineered features. This motivates the inclusion of strategies to automatically extract features from raw sensor data, such as the use of a convolutional neural network (CNN).

The objectives of the current chapter are to: 1) develop a method of encoding raw orientation data such that a CNN could be leveraged for automatic feature extraction in both time and spatial domains of multivariate sensor data and, 2) accurately classify activities of daily living performed by subjects instrumented with only knee wearable sensors with focus on activities influencing satisfaction in patients following total knee replacement.

7.1 Methods

Methods for human activity recognition (HAR) presented in the current chapter will be demonstrated using the system in Chapter 4. Recall, each sensing unit independently estimates its orientation in space with respect to a global reference frame by fusing raw accelerometer, gyroscope, and magnetometer readings.

The instantaneous output of this system is a time-synchronized set of four quaternion rotations, each corresponding to a sensor's orientation. Data are logged at of 25 Hz, which has been shown sufficient to measure the lower extremities [59]. When measured across time, logged quaternions can be observed more generally as a discrete multivariate time series.

Given a single signal sampled at a frequency f for a time of s, the univariate discrete time series U would have the format $U = [t_1, \ t_2, \ ... \ t_{f \times s}]$ where t_x are time-ordered sample values. Given n signals, or a multivariate signal with n dimensions, the corresponding multivariate time series M would have the format $M = [u_1, \ u_2, \ ... \ u_n]$ where each u_x is a univariate discrete time series.

7.1.1   Data Collection

Twenty healthy subjects (age: 24 ± 3 years, M:F=11:9) were instrumented with a previously developed wearable sensor system designed to monitor the legs during activity [88]. All subjects participated voluntarily, informed consent was collected, and ethics approval was obtained from the University of Western Ontario Research and Ethics board prior to testing (Appendix A). Each subject performed 11 activities of daily living (described in Table 7.1) to be classified by machine learned models. A combination of static and dynamic activities were chosen to encompass a majority of daily living. High knee flexion activities selected such as deep/left/right kneeling as well as navigating stairs correspond to influences of satisfaction following joint replacement in Western and non-Western cultures. Subjects executed activities at their own regulated pace and were not instructed to adjust posture or task execution during recording. All data acquisition was performed in the same recreational facility on multiple days in a cardiovascular training area with a treadmill, stair-climbing machine, recumbent bicycle, and floor mats. Approximately thirty seconds of each activity was recorded for each subject and data were manually trimmed to remove any transitions from one activity to another.

7.1.2   Windowing

The discrete multivariate time series of quaternions recorded by the wearable system were first windowed to classify more instantaneous sections of activity. More frequent predictions can be made using a shorter window but available activity context is lowered and accuracy may drop. A window size of 64 samples was chosen which corresponds to approximately 3.2 seconds with a data logging frequency of 25Hz.

Periods of activity (30s) per subject were windowed by shifting only a single sample to

Table 7.1: Description of activities performed by subjects

Activity	Activities performed
Walking	Both slow and fast walking on a treadmill
Running	Both slow and fast running on a treadmill
Stairs	Ascending stair-climbing machine, ascending and descending a staircase
Cycling	Cycling on a stationary machine
Standing	Upright, motionless standing
Sitting	Chair sit with feet on the floor
Single Kneel	Left knee on ground, right knee on ground
Deep Kneel	Both knees on ground with back straight and sitting on upper legs
Laying	Flat against the ground in a sleeping position

maximize the amount of windows per session and to include samples beginning at all stages of the activity performed to contain windows starting and ending at various points along the performed activity. Although a final deployment of the trained classification model could also use highly overlapped windows, it is impractical to perform classifications several times a second, or on every new sample acquired.

These windows are shorter than those used by Kwapisz *et al.* and Voicu *et al.* since Patel *et al.* found optimal windows to between 1s to 4s [138]. The sample size of 64 was chosen as a nearest multiple of eight as is common for CNN image input sizes.

7.1.3 Activity Tiles

Since windowed data are structured as ordered sets of vectors, it was a natural choice to stack these multivariate time series sets to form a matrix. For each time sample from the wearable system, each of the four individual sensor readings consists of four quaternion elements such that a 16 element vector feature set is exhibited. A system with more sensors could be used with the additional sensor readings concatenated. Since it is expected that relations between sensors are important (in addition to the raw readings) for discriminating different recorded activities, windows were augmented with this additional information. This stacking maintains the spatial relevance that exists between the sensors are mounted on a subject, with one above and below each knee.

A quaternion difference between the upper and lower sensors for each leg was computed

(a) Concatenated quaternion components.

(b) Quaternions rearranged so similar components are spatially near.
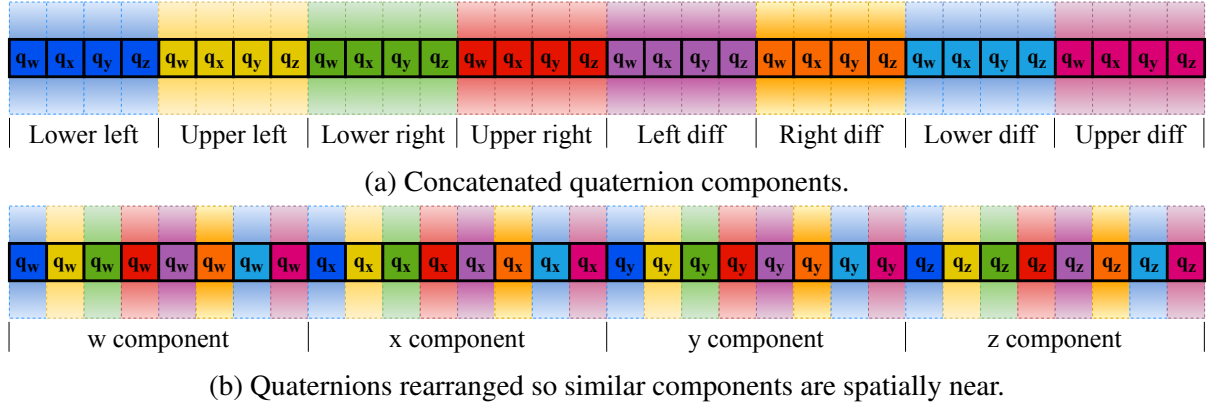
Figure 7.1: Single row of an activity tile formed by horizontally concatenating sensor readings.

using eq. (2.14) and concatenated to the existing 16 elements to form a sample 24 elements wide. Additionally, the rotational difference between both upper sensors and both lower sensors was computed and concatenated, extending the feature set to 32 elements. This $32px \times 64px$ image will further be referred to as an *activity tile*.

It was previously highlighted that spatial relations between image pixels are expected to be important in this classification task. To maximize benefits of automatic feature extraction using a CNN, quaternion rows were rearranged to place contextually similar quaternion components adjacent to one another. The initial activity tile row structure formed through concatenation or each raw sensor reading and their rotational differences can be seen in Figure 7.1a while the transformed version grouping similar quaternion components can be seen in Figure 7.1b. To further increase the dimensionality of activity tiles, grayscale (single channel) pixel values between -1 and 1 were encoded to red-green-blue (RGB) values using a heatmap scheme. A grayscale pixel value P_{Gray} was converted to a three-channel RGB pixel P_{RGB} using the following criteria:

$$R(P_{Gray}) = clamp_{[0,1]}(2 - |2P_{Gray} - 2|)$$
$$G(P_{Gray}) = clamp_{[0,1]}(2 - |2P_{Gray}|)$$
$$B(P_{Gray}) = clamp_{[0,1]}(2 - |2P_{Gray} + 2|)$$

(7.1)

where $clamp_{[0,1]}(x)$ is defined as 0 if $x<0$, x if $0 \le x \le 1$, and 1 if $x>1$. This scheme is similar to the *hot-cold* colour encoding that can be found in Matlab (Mathworks, Natick, USA). An example of an RGB encoded activity tile for several activities can be seen in Figure 7.2.
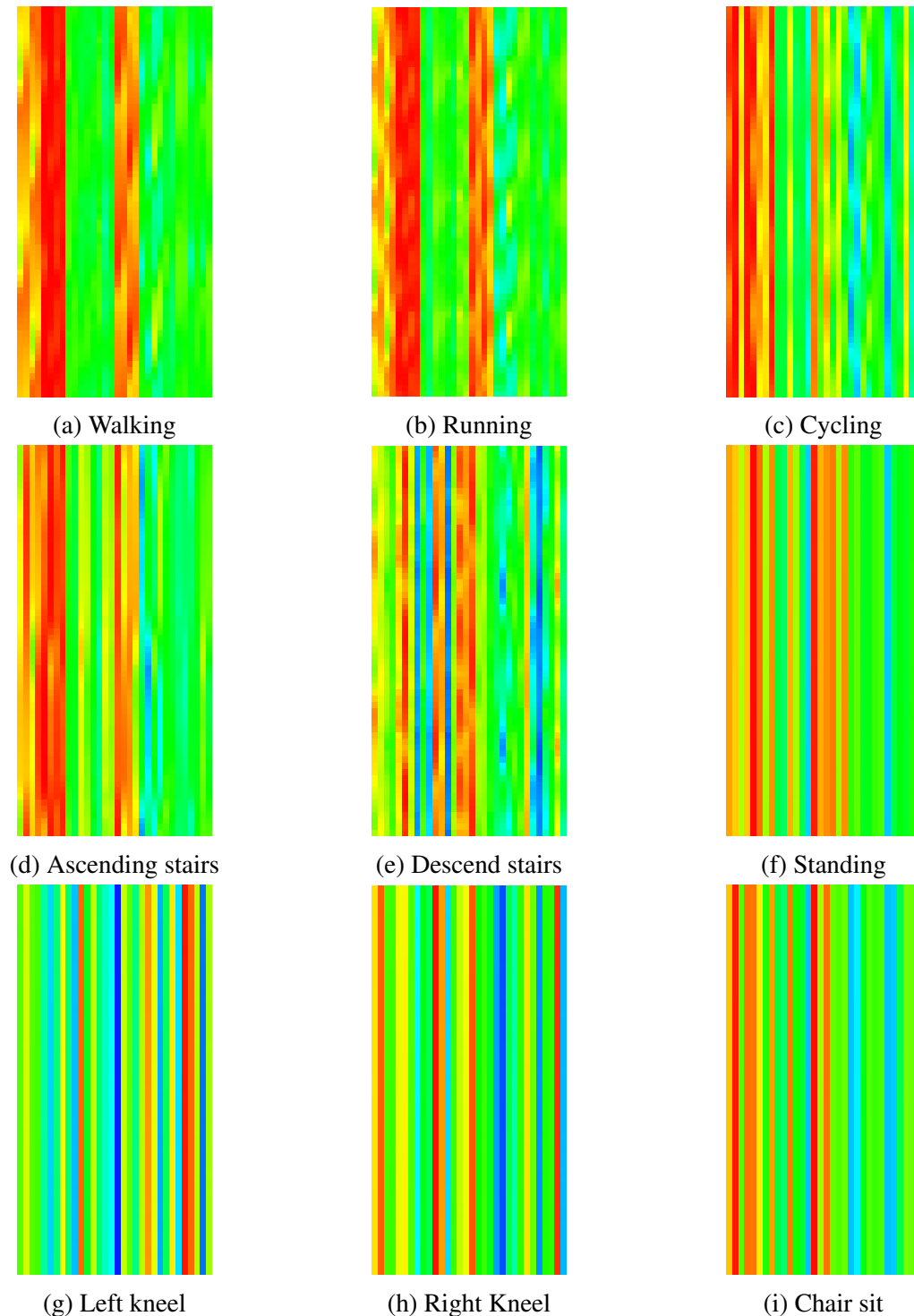
Figure 7.2: Sample activity tiles formed by vertically concatenating rows of grouped quaternion components. The vertical axis of each tile demonstrates changes in sensor readings over time and the horizontal axis is the set of arranged quaternion components from the sensor outputs (Figure 7.1b). Gradients in the vertical axis of dynamic tiles (Figures 7.2a to 7.2e) indicate the subject is moving, which contrasts the static tiles (Figures 7.2f to 7.2i) that remain relatively consistent over time.

This transformation separates strong positive values onto the red channel, neutral values to the green channel, and strong negative values to the blue channel. Although no additional information is added to the images by separating the input channels, it is expected that information distinguishing the activity performed will not be equally dispersed among each channel. This method expands the number of parameters trained early in the network and permits different filters than if the image was input as a single channel.

7.1.4 Data Augmentation

A caveat of measuring orientations with respect to a global frame is the existence of an inherited similarity among quaternions for all activities performed while facing the same compass direction. To remove this similarity, each tile in the data set was duplicated, and duplicates were virtually rotated about the global vertical axis using (2.13). Each tile in the original data was duplicated ten times, with each duplication globally rotated an additional random increment between 0 and 72 deg of vertical axis rotation. By performing these increments, the data set is expanded to include tiles with the same activity profile but as if activities were performed facing alternate compass directions. Varying the rotation between 0 and 72 deg for 10 duplications introduces tiles varying randomly between 0-720 deg from the subject's initial direction.

In addition, to compensate for variations in the placement of sensors on subjects and superficial sensor motion during movement, slight variations in sensor outputs were virtually imposed on the upper limb mounted sensors on both legs. It is expected that soft tissue artifacts will be more prevalent on the upper limbs due to increased soft tissue compared to the lower legs. For each tile, upper sensor readings were locally rotated about their local frame randomly $\pm 10 deg$ along the axis aligned with the upper leg using (2.12). This augmentation is performed for increased generalization between subjects who may demonstrate more variation in upper leg sensor placement due to skin motion artifacts during activities.

By performing both of these augmentations the data set should be more generalized to activities performed by different subjects and inter-subject variability will be minimized.

7.1.5 Network Design

A shallow CNN structure with only three fully convolutional layers was used. This was motivated by the popular *LeNet* model, which has been proven to perform well on relatively small images ($\approx$28x28px) in tasks such as written digit recognition [139]. A similar shallow network was expected to produce sufficient abstraction to differentiate the activity tiles formed. Images input to the network after colour encoding had height of 64px, width of 32px, and depth of three for all RGB colour channels.

To assist in developing a model to classify activities executed at variable speeds and to decrease time variance across samples, a vertical filter with a height of three and width of one was used for the first convolutional layer. Layers with square filter sizes are more common and have the capability of learning filters with responses to both horizontally and vertically oriented features in images. By using a vertical filter, it is expected that the first layer (least abstract) will develop filters responding to features across the time domain only. The height of this filter was chosen to be the smallest odd filter size because of the small input image dimensions. To maintain the granularity of activity features across time, a stride of one was chosen for similar reasons and a *valid* padding scheme was chosen to avoid padding edges with zero values and losing clarity of edge samples.

After the first convolutional layer, a batch normalization layer was added to decrease training time [139]. Following a *ReLu* activation function, an average pooling layer was added with a pool size of 2x2. Average pooling was chosen over max pooling and other alternatives to compliment the organization of quaternion components in activity tiles and to maintain a distinction between two adjacent pixels that may both have high magnitude values and a pair that may only have a single high value. If two adjacent pixels have values $P_1 = 0.05, P_2 = 0.8$, averaging the two values would give a response of 0.43 while the maximum would give 0.8 without consideration of the low value pixel. Features in the activity tiles are expected to be small and localized, so a small pool size and averaging are expected to maintain features. Pooling layers were included to slightly increase the variability of feature locations in activity tiles. Dropout layers were avoided due to the shallow structure of the model.

A second and third convolutional layer were added in the same structure with square 3x3 fil-

ters intended to form further abstraction and respond to relations in the time-domain responses of previous layers. Resulting layers were flattened, fully connected, and passed through a *soft-max* activation function. A summary of the chosen network structure can be seen in Figure 7.3.
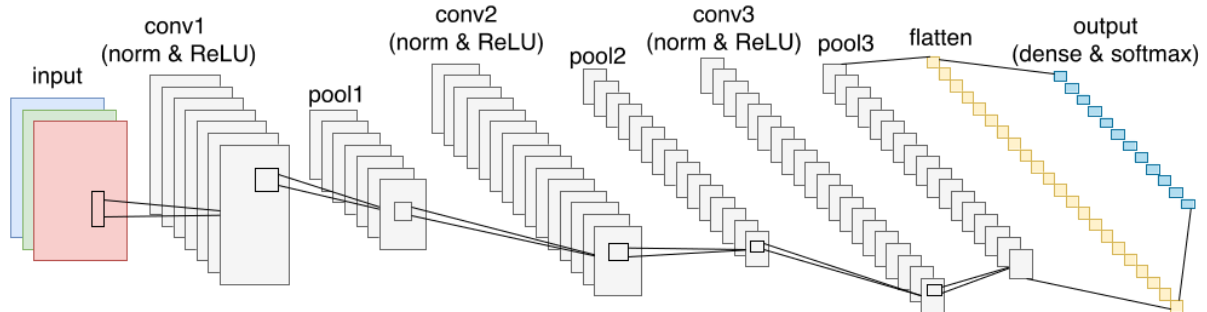
Figure 7.3: A summary of the CNN structure described in Section 7.1.5

7.1.6 Convolutional Neural Network Training and Testing

Activity tiles were saved to disk as PNG image files and were separated into labelled directories using a custom program written in Swift 5.0. A Python (v3.6.8) script was created using Keras (v2.2.4) and Tensorflow (GPU v1.9.0) to train the classification models. All organized and labelled directories were batch-streamed into memory using a Keras library ImageData-Generator object. Samples were shuffled every epoch and the network was permitted to train a maximum of five epochs. Early stopping was implemented to interrupt training when the validation loss did not decrease, and the maximum epoch count was only reached a single time during training. A categorical cross-entropy loss function was employed and models were optimized using the Adam optimizer, which uses a momentum term when adjusting the learning rate and can converge faster than the more traditional stochastic gradient descent [140]. Batches of 256 samples were used to maximize the utility of the Nvidia GTX 1080ti used for training since images were small (32px × 64px) and larger batches could be loaded into the available video memory. On each iteration, a validation split of 0.2 was used to optimize the training parameters. Hyperparameters such as the network depth were tuned using manual selection to get the best performance on both training and validation sets while observing low validation loss to minimize model overfitting.

For testing the developed models, a bash script was used to iteratively create training activ-

ity tiles for 19 subjects and a separate test set from the remaining subject to ensure no samples from the testing patient were contained in the training set. At the end of each iteration, all test and training data were removed from disk storage and activity tiles were re-created using a different test subject. Each time activity tiles were parsed from the raw motion data, randomness was introduced through the global quaternion rotations and upper sensor noise additions so this process is expected to further improve the generalization of these results to future unseen subjects. This process was repeated for the training and testing of four models. Each independent model was trained to perform a different classification in a short hierarchical structure, motivated by previous improvements using ensembles of classifiers to improve accuracy [99].

All classes

A CNN model with the described structure was trained to first distinguish all 11 classes of activities. This classifier will provide a baseline for the other models to evaluate hierarchical structure improvements.

Static vs. Dynamic

To train this model, all samples were sorted into two directories: static and dynamic. It was expected that a primary classifier could be trained with a high degree of accuracy to distinguish static and dynamic activities.

Static

Given a sample is known to be a static activity, this model will be trained to distinguish one of the six classes: sitting in a chair, standing, left/right/deep kneeling, or laying.

Dynamic

If it is known a sample is a dynamic activity, this model will classify it into one of the following five classes: cycling, running, ascending/descending stairs, or walking.

7.1.7 Evaluation

On each iteration of training each model with structure as described above, a LOSO vali-
dation was used. A single subject was left out on each iteration as a test subject. Classification
accuracy was recorded as well as the F-score for each class.

Accuracy and F-score was computed for each activity as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$
$$Precision = \frac{TP}{TP + FP}, \ Recall = \frac{TP}{TP + FN} \quad (7.2)$$
$$F\text{-}Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Through 20 iterations, a mean and standard deviation was used to detect final values inter-
subject differences.

7.2 Results

Activities of daily living (Table 7.1) were performed by all subjects but during data pro-
cessing it was found that the body-worn sensors had slipped substantially on the legs of two
subjects during the cycling activity and running from one subject. These trials were withheld
from model training and evaluation since it was visibly obvious that data were not correct and
these distorted samples would be expected to introduce error into developed models.

Classification accuracy performed by a single model for all activities and the F-score for
each individual activity can be seen in Table 7.2. Activities with the lowest F-score are cycling
(0.72), ascending (0.79), and descending (0.68) stairs. It can also be seen that static tasks in
general were classified more successfully than the dynamic tasks with the lowest classification
accuracy on the chair sit activity (0.90).

A second model to distinguish static activities from dynamic was evaluated and results can
be seen in Table 7.3. This classifier was able to almost perfectly distinguish tasks (acc.: 1.00)
involving motion from those that are stationary.

Finally, given the success of Model 2 to distinguish static and dynamic activities, the two
additional models were evaluated separately to distinguish individual static and dynamic activ-

Table 7.2: Classification of all activities in a single model: total accuracy for all tasks and F-score for each activity.

	Mean	Std Dev		Mean	Std Dev
Accuracy	**0.91**	**0.10**	Right Kneel	1.00	0.00
Cycle	0.72	0.40	Run	0.93	0.11
Chair Sit	0.90	0.30	Stair Ascend	0.79	0.30
Deep Kneel	1.00	0.00	Stair Descend	0.68	0.33
Laying	1.00	0.00	Stand	0.99	0.02
Left Kneel	0.94	0.18	Walk	0.93	0.19

Table 7.3: Static vs. Dynamic Activities

	Accuracy	Static F	Dynamic F
Mean	**1.00**	1.00	1.00
Std. Dev	0.01	0.01	0.01

ities (Table 7.4). It was found that the accuracy of classifying static activities was nearly perfect (acc.: 0.99) with the most confused activity being the chair sit task (acc.: 0.97). Low standard deviation values when evaluating across all subjects indicates consistently high accuracy was observed and it would be expected that similar results would be observed when deployed on unseen future subjects.

Table 7.4: Mean total classification accuracy for all activities and F-score of each activity for two independent models, one for static and the other for dynamic tasks.

Static			Dynamic		
	Mean	Std Dev		Mean	Std Dev
Accuracy	**0.99**	**0.03**	**Accuracy**	**0.91**	**0.09**
Chair Sit	0.97	0.12	Cycle	0.84	0.24
Deep Kneel	0.98	0.07	Run	0.98	0.04
Laying	1.00	0.00	Stair Ascend	0.86	0.25
Left Kneel	0.99	0.02	Stair Descend	0.78	0.22
Right Kneel	0.99	0.03	Walk	0.95	0.10
Standing	0.99	0.02			

Dynamic classification accuracy remained similar to the combined model however, F-scores improved for all static tasks. Activities classified with the lowest F-score in the previous combined model have improved (cycle: 0.84, ascending: 0.86, descend: 0.78) but still remain the lowest of all tasks.

Two confusion matrices were generated by averaging the mean classification success rate for each activity over all iterations. The confusion matrix for static activities is seen in Figure 7.4a while the dynamic is shown in Figure 7.4b.

(a) Static   (b) Dynamic
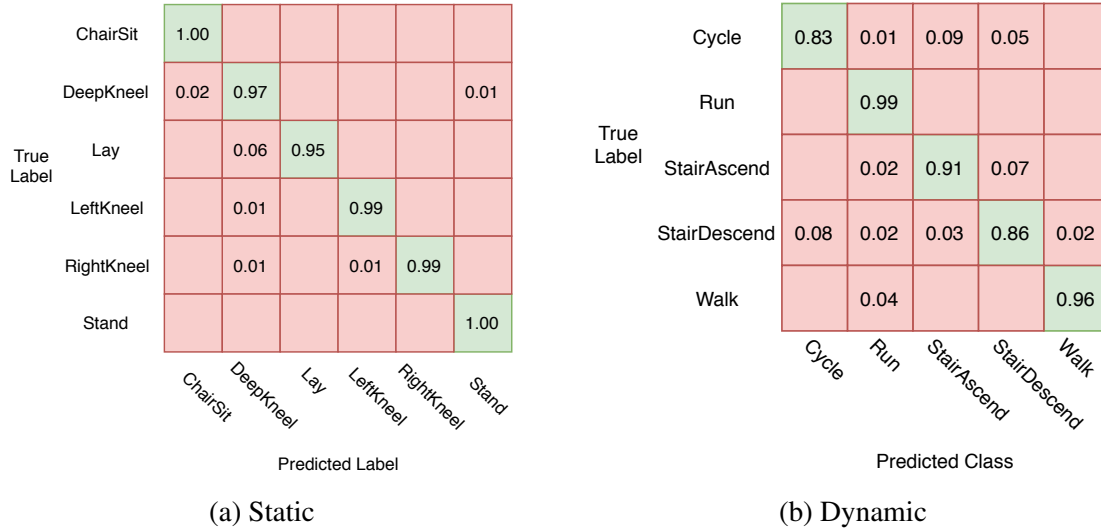
Figure 7.4: Confusion matrices for static and dynamic classification models.

Few static activities were misclassified and no definite pattern presented across all iterations, however, it was evident from Figure 7.4b that cycling (acc. 0.83) and descending stairs (acc. 0.86) were more frequently misclassified. A large portion (7%) of stair ascend activities were also classified as stair descend. Some walking samples (4%) were incorrectly predicted to be running.

7.3   Discussion

Large standard deviations in F-scores for dynamic activities shown in Table 7.2 and Table 7.4 suggest that there were inter-subject differences in task execution and classification was more successful for some subjects than others. This was especially true for cycling and ascending/descending stairs, which were most commonly confused in the current work as well as previous publications [93, 98]. Future work could benefit from incorporating an altimeter which could potentially distinguish these activities based on an increasing, decreasing, or constant altitudes. Future work in this area could also introduce a calibration step where subjects perform each activity once to be included in the training set before performing additional in-

(a) Walking   (b) Running

Figure 7.5: Confused walking and running samples for a single subject.

strumentation sessions. In a deployment of this system to remotely monitor subjects, this is a technique that could be used to minimize the inter-subject difference in movements.
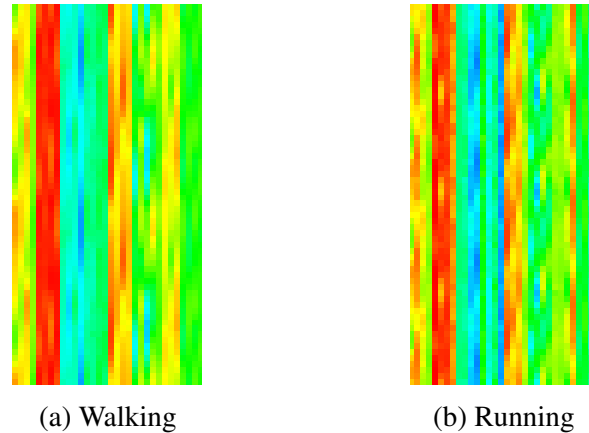
It should be noted that execution speed was not dictated when subjects performed the specified activities. Subjects performed tasks at their own desired pace so the data recorded vary in execution speeds. This was a conscious decision since it was desired to evaluate the work presented in realistic deployment scenarios. A consequence of this variation in pace was an increase in confused samples and classifications, seen most prominently for faster-moving activities Figure 7.4b. On one iteration of validation for the dynamic activity classification model, one subject's *walking* samples were partially classified as *running*. It is expected that this subject simply walked at a faster pace that may have been close to the pace of some subjects when running. A *walking* activity tile of this subject compared to a *running* tile from another subject in the training set can be seen in Figure 7.5. Visual comparison of these two tiles reveals the similarity in samples.

Recall the wearable sensor system deployed recorded quaternion components at a frequency of 25Hz and trained networks used a filter size of 3x1 on the first convolutional layer to accentuate features across the time domain only. If results were to be replicated using a similar data structure recorded at a higher frequency, it is expected that a more rectangular filter, such as 5x1, may be needed to proportionally respond to more data points of a similar length of activity. Similarly, it is expected that increasing the height of formulated activity tiles to encompass the same length of time would be beneficial for reproducing similar results.

The current work leveraged a CNN to detect static vs. dynamic activities but this stage could be replaced by thresholding motion variations. It was noted that subjects shifted their weight and moved even during static activities so it is expected that the current method would provide robust distinctions even when activities are not completely motionless.

Activities of subjects in the current work were stripped of all transition periods moving from a one activity to another. Subjects were first asked to begin the specified activity and recording was started by an observer. It is expected that accuracy of the model presented would drop if transitions were included in the study since it was observed that subjects exhibited substantially different strategies for their transitions.

Unfortunately three activities between two subjects were removed due to sensor straps slipping on the users. It is expected that this was caused by mounting the sensors using the hook-and-latch straps on the outside of fitness-type, loose-fitting material and the faster movements of cycling and running caused the sensors to slip and record incorrect and un-classifiable data. Future development to the wearable sensor system may be required to instrument patients while performing fast-motion activities to prevent incorrect classifications.

Although the current work has shown the benefits of implementing the deployed four-sensor wearable system for the improvement of activity classification, further investigation into subject coherence with wearing the system outside of observation must be explored. Systems with single sensor components are more convenient to attach/reattach to body segments. Some systems using only an accelerometer are intended to be worn for very long periods of time (≈ 30 days) and can operate on a low-energy draw. Based on observations of the deployed sensor system in this study, recording periods of this length will not be possible while maintaining a logging rate of 25Hz on each sensor. Further refinements to the current hardware could be made to improve battery life or subjects could dock the sensors in a charging station overnight to recharge batteries for more extended wear.

7.4 Contribution

The current work encoding raw orientation data as images and training a CNN for classi-fications has shown high accuracy for daily living activities. Distinctions can be very confi-

dently made between static and dynamic activities and further refinement of the activities was also demonstrated. This study has shown that minimal instrumentation of a subject's knees can provide a strong distinction of daily activities. Although a fixed number of activities have been used for this classification task, it is expected that these could include the majority of activities performed daily. The content and findings of this chapter have been published in the journal *IEEE Sensors* [141].

Chapter 8

Conclusions and Future Work

The final remaining chapter of this thesis will summarize the contributions made in the current work and discuss several areas of future work that could be further improved and expanded with continuous research and development. This work presented in this thesis was aimed to develop an appropriate tool to functionally instrument patients in the clinic. Objective measurement is crucial for the early detection of joint performance issues as well as properly assessing patients. Unfortunately due to limited time in the clinic, extensive functional assessment is rarely performed and surgeons and clinicians must rely on subjective patient feedback for evaluation. It was shown that functional measurements should be incorporated into pre-surgical evaluations and to accompany self-reported measures at clinical follow-ups to help address traditional outcome evaluation shortfalls. The developed measurement tool has been shown appropriate for use in the clinic following hundreds of instrumentation sessions with many patients both before and after undergoing total knee replacement.

8.1 Contributions

The specific contributions of this thesis can be summarized as the following:

- The development of an instrumentation platform that can be easily deployed in the clinic to obtain granular, objective functional measurements from patients with knee osteoarthritis undergoing total knee replacement. The system has been validated using

a robot phantom and a gold standard motion capture measurement system to ensure the accuracy of joint angle measurements with visual sensor replacement. In addition to this validation, the continued use of the system in the orthopaedic clinic at London's University Hospital through several research studies has shown continuous utility and practicality of the tool.

- In order to measure patient function repeatably and appropriately for cross-patient comparisons, a novel instrumentation and measurement strategy was developed to obtain useful joint performance metrics from the timed-up-and-go (TUG) performance test. Although previous related work has instrumented this test with wearable sensors, no knee specific tools were practical for deployment in a time-constrained clinical setting to accomplish this task. A novel test segmentation strategy was presented using the orientations of the upper and lower segments of both legs to separate the TUG into distinct joint-stressing activities. From these segments, joint metrics were specifically to measure knee joint function. By instrumenting knee patients one year post surgery, these metrics revealed some characteristics of joint performance common to patients who are satisfied or dissatisfied following their surgery.

- Using TUG tests from preoperative patient visits, unsupervised machine learning was used to determine that specific functional metrics could help predict if patients were likely to experience measurable functional improvement following surgery. This advancement in the field of knee outcomes was highlighted by the Arthritis Society as a top 10 advancement of the year. Unmet patient expectations following surgery is one of the largest risk factors for patient dissatisfaction; a persisting problem experienced by up to 1 in 5 patients. Cluster analysis revealed two groups of patients separated only using derived functional metrics and each group experienced different recovery paths with recovery responders generally having worse function preoperation but demonstrating nearly all of their recovery as early as three months post-surgery.

- Following the discovery of preoperative patient function indicating likelihood of functional recovery, classifier models were explored to determine if a patient could be individually predicted to become a recovery responder or maintainer before undergoing

surgery. It is expected that given this information, the surgeon could help realistically adjust patient expectations. It was found that accurate and highly sensitive models could be trained using only functional metrics as features. Model performance was improved by adding additional patient information such as demographics and self-reported measures, however, obtaining this additional information is at the expense of time and effort in the clinic.

- In the final chapter, a novel algorithm for performing human activity recognition was developed. A convolutional network was leveraged by encoding raw orientation measurements into images to perform automatic feature extraction. It was found that very accurate models could be trained to recognize activities of daily living performed by healthy subjects in a fitness centre setting. This work presented promising results for future deployments of this sensor system for at-home monitoring free of observer bias.

8.2 Future Work

During the compilation of the completion of contained work, several areas for future work and continuation have been noted:

- Skin motion artifacts have been noted as an issue with any surface-mounted wearable sensor system. Further investigation into modelling skin motion could reveal compensation strategies to help relate measured leg segment orientations more closely to the femur and tibia orientation less influenced by skin motion artifacts. One possible avenue is the incorporation of a machine learning model to learn motion compensation if true ground truth measurements can be obtained free of skin motion.

- In the current work a fixed set of functional metrics have been derived from the TUG test. Although they have shown to be effective for predicting recovery and assessing function, they are by no means expected to exhaustively represent function. Additional, and perhaps, more useful metrics could still be derived from raw sensor data. As highlighted in Chapter 7, automatic feature extraction can often out-perform solutions using

manually engineered features in classification tasks. As more patient data becomes available, it may be possible to apply similar automatic extraction to entire sessions of TUG recordings opposed to the short activity samples explored in the current work.

- In addition to further functional metrics, more information is collected clinically at a follow-up than is currently being used. It may be that a model using additional parameters such as inflammatory markers, implant size and information, surgical technique and even details extracted from preoperative clinical radiographs could reveal much more about patient status' for more individualized care suggestions and predictions.

- Lastly, this thesis has touched on remote instrumentation through human activity recognition. Currently more research must take place to determine the effect of daily activity performance compared to in-clinic demonstrated performance. Recently visiting the clinic for post-surgery follow-ups has been difficult or impossible for many patients and the need for remote technologies to help surgeons assess their patients has been highlighted. Further work should assess the ability for patients to use this system without an observer for more frequent instrumentation to develop a clearer picture of their recovery path, and possibly alert surgeons of possible problems.

Bibliography

[1] Canadian Institute for Health Information, "Hip and Knee Replacements in Canada, 2017–2018: Canadian Joint Replacement Registry Annual Report," Canadian Institute for Health Information, Ottawa, Tech. Rep., 2019.

[2] B. R. Deshpande *et al.*, "Number of Persons With Symptomatic Knee Osteoarthritis in the US: Impact of Race and Ethnicity, Age, Sex, and Obesity," *Arthritis Care and Research*, vol. 68, no. 12, pp. 1743–1750, 2016.

[3] A. J. Price *et al.*, "Knee replacement," *Lancet*, vol. 392, no. 10158, pp. 1672–1682, 2018.

[4] R. B. Bourne *et al.*, "Patient satisfaction after total knee arthroplasty: Who is satisfied and who is not?" in *Clinical Orthopaedics and Related Research*, vol. 468, no. 1. Springer New York, 2010, pp. 57–63.

[5] M. J. Dunbar, G. Richardson, and O. Robertsson, "I can't get no satisfaction after my total knee replacement," *The Bone and Joint Journal*, vol. 95-B, no. 11_Supple_A, pp. 148–152, 2013.

[6] R. Gunaratne *et al.*, "Patient Dissatisfaction Following Total Knee Arthroplasty: A Systematic Review of the Literature," 2017.

[7] A. Peres-Da-Silva *et al.*, "What Factors Drive Inpatient Satisfaction After Knee Arthroplasty?" 2017.

[8] C. Tilbury *et al.*, "Primary Arthroplasty Unfulfilled Expectations After Total Hip and Knee Arthroplasty Surgery: There Is a Need for Better Preoperative Patient Information and Education," 2016.

[9] I. Press and F. Fullam, "Patient satisfaction in pay for performance programs," *Quality Management in Health Care*, vol. 20, no. 2, pp. 110–115, 2011.

[10] P. Chatterjee *et al.*, "Patient experience in safety-net hospitals: Implications for improving care and value-based purchasing," *Archives of Internal Medicine*, vol. 172, no. 16, pp. 1204–1210, 2012.

[11] J. L. Berliner *et al.*, "Can Preoperative Patient-reported Outcome Measures Be Used to Predict Meaningful Improvement in Function After TKA?" *Clinical Orthopaedics and Related Research®*, vol. 475, no. 1, pp. 149–157, 2017.

[12] H. Y. Yang *et al.*, "Longitudinal Trajectories of Pain and Function Improvement Following Total Knee Replacement," *ACR Open Rheumatology*, vol. 1, no. 5, pp. 308–317, 2019.

[13] C. A. Kahlenberg *et al.*, "Patient Satisfaction After Total Knee Replacement: A Systematic Review," *HSS Journal* ®, vol. 14, no. 2, pp. 192–201, 2018.

[14] A. L. Miner *et al.*, "Knee range of motion after total knee arthroplasty: How important is this as an outcome measure?" *The Journal of Arthroplasty*, vol. 18, no. 3, pp. 286–294, 2003.

[15] B. L. Conner-Spady *et al.*, "Patient expectations and satisfaction 6 and 12 months following total hip and knee replacement," *Quality of Life Research*, vol. 29, no. 3, pp. 705–719, 2019.

[16] M. P. Siljander *et al.*, "Current Trends in Patient-Reported Outcome Measures in Total Joint Arthroplasty: A Study of 4 Major Orthopaedic Journals," *Journal of Arthroplasty*, vol. 33, no. 11, pp. 3416–3421, 2018.

[17] C. W. Grayson *et al.*, "Revision Arthroplasty Functional Improvement and Expectations Are Diminished in Total Knee Arthroplasty Patients Revised for Flexion Instability Compared to Aseptic Loosening and Infection," 2016.

[18] P. J. C. Heesterbeek *et al.*, "Moderate clinical improvement after revision arthroplasty of the severely stiff knee," *Knee Surg Sports Traumatol Arthrosc*, vol. 24, pp. 3235–3241, 2016.

[19] J. Moya-Angeler *et al.*, "Revision Arthroplasty Revision Arthroplasty for the Management of Stiffness After Primary TKA," 2017.

[20] A. Rajgopal *et al.*, "Revision Arthroplasty Are the Outcomes of Revision Knee Arthroplasty for Flexion Instability the Same as for Other Major Failure Mechanisms?" 2017.

[21] M. A. Tully, J. Panter, and D. Ogilvie, "Individual Characteristics Associated with Mismatches between Self-Reported and Accelerometer-Measured Physical Activity," *PLOS ONE*, vol. 9, no. 6, pp. 1–9, 2014.

[22] I. E. Luna *et al.*, "Early patient-reported outcomes versus objective function after total hip and knee arthroplasty," *Bone & Joint Journal*, vol. 99-B, no. 9, pp. 1167–1175, 2017.

[23] R. G. Marx *et al.*, "Measuring improvement following total hip and knee arthroplasty using patient-based measures of outcome," *Journal of Bone and Joint Surgery - Series A*, vol. 87, no. 9 I, pp. 1999–2005, 2005.

[24] S. A. A. N. Bolink, B. Grimm, and I. C. Heyligers, "Patient-reported outcome measures versus inertial performance-based outcome measures: A prospective study in patients undergoing primary total knee arthroplasty," *Knee*, vol. 22, no. 6, pp. 618–623, 2015.

[25] C. Malchow and G. Fiedler, "Effect of observation on lower limb prosthesis gait biomechanics: Preliminary results," *Prosthetics and Orthotics International*, vol. 40, no. 6, pp. 739–743, 2016.

[26] A. T. Berman *et al.*, "Quantitative gait analysis after unilateral or bilateral total knee replacement," *Journal of Bone and Joint Surgery - Series A*, vol. 69, no. 9, pp. 1340–1345, 1987.

[27] N. Arden and M. C. Nevitt, "Osteoarthritis: Epidemiology," pp. 3–25, 2006.

[28] T. D. Spector and A. J. MacGregor, "Risk factors for osteoarthritis: Genetics," *Osteoarthritis and Cartilage*, vol. 12, no. SUPLL., pp. 39–44, 2004.

[29] G. Peat *et al.*, "Clinical classification criteria for knee osteoarthritis: performance in the general population and primary care," *Ann Rheum Dis*, vol. 65, pp. 1363–1367, 2006.

[30] C. Palazzo *et al.*, "Risk factors and burden of osteoarthritis," pp. 134–138, 2016.

[31] T. Vos *et al.*, "Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010," Tech. Rep., 2012.

[32] M. Cross *et al.*, "The global burden of hip and knee osteoarthritis: estimates from the Global Burden of Disease 2010 study," 2014.

[33] M. Cucchiarini *et al.*, "Basic science of osteoarthritis," 2016.

[34] J. A. Wallis *et al.*, "What proportion of people with hip and knee osteoarthritis meet physical activity guidelines? A systematic review and meta-analysis," pp. 1648–1659, 2013.

[35] B. Heidari, "Knee osteoarthritis prevalence, risk factors, pathogenesis and features: Part I," *Caspian Journal of Internal Medicine*, vol. 2, no. 2, pp. 205–212, 2011.

[36] B. Sharif *et al.*, "Productivity costs of work loss associated with osteoarthritis in Canada from 2010 to 2031," *Osteoarthritis and Cartilage*, vol. 25, no. 2, pp. 249–258, 2017.

[37] T. E. McAlindon *et al.*, "OARSI guidelines for the non-surgical management of knee osteoarthritis."

[38] L. M. Billesberger *et al.*, "Procedural Treatments for Knee Osteoarthritis: A Review of Current Injectable Therapies," 2020.

[39] C. J. Vertullo *et al.*, "Surgeon's Preference in Total Knee Replacement: A Quantitative Examination of Attributes, Reasons for Alteration, and Barriers to Change," *Journal of Arthroplasty*, vol. 32, no. 10, pp. 2980–2989, 2017.

[40] B. K. Daines and D. A. Dennis, "Gap balancing vs. measured resection technique in total knee arthroplasty," pp. 1–8, 2014.

[41] N. Bellamy *et al.*, "Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee." *Journal of Rheumatology*, vol. 15, no. 12, pp. 1833–1840, 1988.

[42] N. Bellamy *et al.*, "Osteoarthritis Index delivered by mobile phone {(m-WOMAC)} is valid, reliable, and responsive," *J Clin Epidemiol*, vol. 64, no. 2, pp. 182–190, 2011.

[43] H. Behrend *et al.*, "The {"Forgotten} Joint" as the Ultimate Goal in Joint Arthroplasty Validation of a New {Patient-Reported} Outcome Measure," *J Arthroplast*, vol. 27, no. 3, pp. 430–436.e1, 2012.

[44] P. W. Stratford, D. M. Kennedy, and D. L. Riddle, "New study design evaluated the validity of measures to assess change after hip or knee arthroplasty," *Journal of Clinical Epidemiology*, vol. 62, no. 3, pp. 347–352, 2009.

[45] P. B. Shull *et al.*, "Quantified self and human movement: A review on the clinical impact of wearable sensing and feedback for gait analysis and intervention."

[46] S. Yao and Y. Zhu, "Wearable multifunctional sensors using printed stretchable conductors made of silver nanowires †," 2014.

[47] T. Q. Trung and N. E. Lee, "Flexible and Stretchable Physical Sensor Integrated Platforms for Wearable Human-Activity Monitoringand Personal Healthcare," *Advanced Materials*, vol. 28, no. 22, pp. 4338–4372, 2016.

[48] A. Muro-de-la Herran, B. García-Zapirain, and A. Méndez-Zorrilla, "Gait analysis methods: An overview of wearable and non-wearable systems, highlighting clinical applications," pp. 3362–3394, 2014.

[49] N. Carbonaro *et al.*, "Exploiting wearable goniometer technology for motion sensing gloves," *IEEE Journal of Biomedical and Health Informatics*, vol. 18, no. 6, pp. 1788–1795, 2014.

[50] J. Wu *et al.*, "Fast Complementary Filter for Attitude Estimation Using Low-Cost MARG Sensors," *IEEE Sensors Journal*, vol. 16, no. 18, pp. 6997–7007, 2016.

[51] R. G. Valenti, I. Dryanovski, and J. Xiao, "A linear Kalman filter for MARG orientation estimation using the algebraic quaternion algorithm," *IEEE Transactions on Instrumentation and Measurement*, vol. 65, no. 2, pp. 467–481, 2016.

[52] L. Zhao and Q. Y. Wang, "Design of an Attitude and Heading Reference System Based on Distributed Filtering for Small UAV," *Mathematical Problems in Engineering*, vol. 2013, 2013.

[53] E. S. Grood and W. J. Suntay, "A joint coordinate system for the clinical description of three-dimensional motions: Application to the knee," *Journal of Biomechanical Engineering*, vol. 105, no. 2, pp. 136–144, 1983.

[54] D. Dabirrahmani and M. Hogg, "Modification of the Grood and Suntay Joint Coordinate System equations for knee joint flexion," *Medical Engineering and Physics*, vol. 39, pp. 113–116, 2017.

[55] F. Dobson *et al.*, "OARSI recommended performance-based tests to assess physical function in people diagnosed with hip or knee osteoarthritis," *Osteoarthritis and Cartilage*, vol. 21, no. 8, pp. 1042–1052, 2013.

[56] S. Van Onsem *et al.*, "Improved walking distance and range of motion predict patient satisfaction after TKA," *Knee Surgery, Sports Traumatology, Arthroscopy*, vol. 26, no. 11, pp. 3272–3279, 2018.

[57] K. Kourou *et al.*, "Machine learning applications in cancer prognosis and prediction," pp. 8–17, 2015.

[58] T. A. Almeida, J. Almeida, and A. Yamakami, "Spam filtering: How the dimensionality reduction affects the accuracy of Naive Bayes classifiers," *Journal of Internet Services and Applications*, vol. 1, no. 3, pp. 183–200, 2011.

[59] M. Sliepen, M. Brandes, and D. Rosenbaum, "Current physical activity monitors in hip and knee osteoarthritis: a review," *Arthritis Care Res (Hoboken)*, vol. 69, no. 10, pp. 1460–1466, 2017.

[60] B. R. Greene, S. J. Redmond, and B. Caulfield, "Fall Risk Assessment Through Automatic Combination of Clinical Fall Risk Factors and Body-Worn Sensor Data," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 3, pp. 725–731, 2017.

[61] G. Cooper *et al.*, "Inertial sensor-based knee flexion/extension angle estimation," *Journal of Biomechanics*, vol. 42, no. 16, pp. 2678–2685, 2009.

[62] J. Favre *et al.*, "Functional calibration procedure for 3D knee joint angle description using inertial sensors," *Journal of Biomechanics*, vol. 42, no. 14, pp. 2330–2335, 2009.

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

[64] H. Dejnabadi, B. M. Jolles, and K. Aminian, "A new approach to accurate measurement of uniaxial joint angles based on a combination of accelerometers and gyroscopes," *IEEE Transactions on Biomedical Engineering*, vol. 52, no. 8, pp. 1478–1484, 2005.

[65] H. Dejnabadi *et al.*, "Estimation and visualization of sagittal kinematics of lower limbs orientation using body-fixed sensors," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 7, pp. 1385–1393, 2006.

[66] F. Alonge *et al.*, "The Use of Accelerometers and Gyroscopes to Estimate Hip and Knee Angles on Gait Analysis," *Sensors*, vol. 14, pp. 8430–8446, 2014.

[67] A. M. Sabatini, "Estimating Three-Dimensional Orientation of Human Body Parts by Inertial/Magnetic Sensing," *Sensors*, vol. 11, no. 2, pp. 1489–1525, 2011.

[68] 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.

[69] F. Dobson *et al.*, "{OARSI} recommended performance-based tests to assess physical function in people diagnosed with hip or knee osteoarthritis," *Osteoarthritis and Cartilage*, vol. 21, no. 8, pp. 1042–1052, 2013.

[70] A. Alghadir, S. Anwer, and J. M. Brismée, "The reliability and minimal detectable change of Timed Up and Go test in individuals with grade 1 - 3 knee osteoarthritis," *BMC Musculoskeletal Disorders*, vol. 16, no. 1, 2015.

[71] E. Yuksel *et al.*, "Assessing Minimal Detectable Changes and Test-Retest Reliability of the Timed Up and Go Test and the 2-Minute Walk Test in Patients With Total Knee Arthroplasty," *The Journal of Arthroplasty*, vol. 32, no. 2, pp. 426–430, 2017.

[72] S. Poitras *et al.*, "Predicting hospital length of stay and short-term function after hip or knee arthroplasty: are both performance and comorbidity measures useful?" *International Orthopaedics*, vol. 42, no. 10, pp. 2295–2300, 2018.

[73] L. J. Viccaro, S. Perera, and S. A. Studenski, "Is timed up and go better than gait speed in predicting health, function, and falls in older adults?" *Journal of the American Geriatrics Society*, vol. 59, no. 5, pp. 887–892, 2011.

[74] B. Mariani *et al.*, "On-Shoe Wearable Sensors for Gait and Turning Assessment of Patients With Parkinson's Disease," {*IEEE*} *Trans Biomed Eng*, vol. 60, no. 1, pp. 155–158, 2013.

[75] B. R. Greene, S. J. Redmond, and B. Caulfield, "Fall risk assessment through automatic combination of clinical fall risk factors and body-worn sensor data," {*IEEE*} *J. Biomed. Health Inform.*, vol. PP, no. 99, p. 1, 2016.

[76] A. Salarian *et al.*, "ITUG, a sensitive and reliable measure of mobility," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 18, no. 3, pp. 303–310, 2010.

[77] Y. Higashi *et al.*, "Quantitative evaluation of movement using the timed up-and-go test," *IEEE Engineering in Medicine and Biology Magazine*, vol. 27, no. 4, pp. 38–46, 2008.

[78] H. P. Nguyen *et al.*, "Auto detection and segmentation of physical activities during a Timed-Up-and-Go (TUG) task in healthy older adults using multiple inertial sensors," *Journal of NeuroEngineering and Rehabilitation*, vol. 12, no. 1, 2015.

[79] P. Ortega-Bastidas *et al.*, "Use of a single wireless IMU for the segmentation and automatic analysis of activities performed in the 3-m timed up & go test," *Sensors (Switzerland)*, vol. 19, no. 7, 2019.

[80] M. El-Gohary *et al.*, "Continuous monitoring of turning in patients with movement disability," *Sensors (Switzerland)*, vol. 14, no. 1, pp. 356–369, 2014.

[81] M. G. Pagé *et al.*, "Distinguishing problematic from nonproblematic postsurgical pain: A pain trajectory analysis after total knee arthroplasty," *Pain*, vol. 156, no. 3, pp. 460–468, 2015.

[82] A. D. Liddle *et al.*, "Determinants of revision and functional outcome following uni-compartmental knee replacement," *Osteoarthritis and cartilage / OARS, Osteoarthritis Research Society*, vol. 22, no. 9, pp. 1241–1250, 2014.

[83] P. R. Fortin *et al.*, "Outcomes of total hip and knee replacement: Preoperative functional status predicts outcomes at six months after surgery," *Arthritis & Rheumatism*, vol. 42, no. 8, pp. 1722–1728, 1999.

[84] P. Thottakkara *et al.*, "Application of machine learning techniques to high-dimensional clinical data to forecast postoperative complications," *PLoS ONE*, vol. 11, no. 5, 2016.

[85] K. N. Kunze *et al.*, "Development of Machine Learning Algorithms to Predict Patient Dissatisfaction After Primary Total Knee Arthroplasty," *Journal of Arthroplasty*, vol. 35, no. 11, pp. 3117–3122, 2020.

[86] F. Kluge *et al.*, "Pre-operative sensor-based gait parameters predict functional outcome after total knee arthroplasty," *Gait and Posture*, vol. 66, pp. 194–200, 2018.

[87] E. Papi, A. Belsi, and A. H. McGregor, "A knee monitoring device and the preferences of patients living with osteoarthritis: a qualitative study," *BMJ Open*, vol. 5, no. 9, 2015.

[88] R. A. Bloomfield *et al.*, "Proposal and validation of a knee measurement system for patients with osteoarthritis," *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 2, pp. 319–326, 2019.

[89] S. Zhang *et al.*, "Physical activity classification using the {G}{E}{N}{E}{A} wrist-worn accelerometer," *Med Sci Sports Exerc*, vol. 44, no. 4, pp. 742–748, 2012.

[90] J. Sasaki, D. John, and P. Freedson", "Validation and comparison of ActiGraph activity monitors," *J Sci Med Sport*, vol. 14, no. 5, pp. 411–416, 2011.

[91] K. Lyden *et al.*, "The activPALTM Accurately Classifies Activity Intensity Categories in Healthy Adults." *Medicine and science in sports and exercise*, vol. 49, no. 5, pp. 1022–1028, 2017.

[92] A. Godfrey *et al.*, "Activity classification using a single chest mounted tri-axial ac-celerometer," *Med Eng Phys*, vol. 33, no. 9, pp. 1127–1135, 2011.

[93] J. Kwapisz, G. Weiss, and S. Moore, "Activity Recognition Using Cell Phone Ac-celerometers," *SIGKDD Explor Newsl*, vol. 12, no. 2, pp. 74–82, 2011.

[94] R.-A. Voicu *et al.*, "Human Physical Activity Recognition Using Smartphone Sensors."

[95] X. Su, H. Tong, and P. Ji, "Activity recognition with smartphone sensors," *Tsinghua Science and Technology*, vol. 19, no. 3, pp. 235–249, 2014.

[96] D. Trong Bui, N. Nguyen, and G.-M. Jeong, "A Robust Step Detection Algorithm and Walking Distance Estimation Based on Daily Wrist Activity Recognition Using a Smart Band," *Sensors*, vol. 18, no. 7, p. 2034, 2018.

[97] D. M. Burns *et al.*, "Shoulder physiotherapy exercise recognition: machine learning the inertial signals from a smartwatch." *Physiological measurement*, vol. 39, no. 7, p. 75007, 2018.

[98] S. de Groot and M. Nieuwenhuizen, "Validity and reliability of measuring activities, movement intensity and energy expenditure with the DynaPort MoveMonitor," *Med Eng Phys*, vol. 35, no. 10, pp. 1499–1505, 2013.

[99] H. Fawaz *et al.*, "Deep learning for time series classification: a review," *Data Min Knowl Disc*, vol. 33, no. 4, pp. 917–963, 2019.

[100] A. Bagnall *et al.*, "The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances," *Data Min Knowl Disc*, vol. 31, no. 3, pp. 606–660, 2017.

[101] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, p. 436, 2015.

[102] J. Bo Yang *et al.*, "Deep Convolutional Neural Networks On Multichannel Time Series For Human Activity Recognition," Tech. Rep.

[103] Z. Wang and T. Oates, "Encoding Time Series as Images for Visual Inspection and Classification Using Tiled Convolutional Neural Networks," Tech. Rep., 2015.

[104] M. Sliepen, M. Brandes, and D. Rosenbaum, "Current Physical Activity Monitors in Hip and Knee Osteoarthritis: A Review," pp. 1460–1466, 2017.

[105] M. P. Kadaba, H. K. Ramakrishnan, and M. E. Wootten, "Measurement of lower extremity kinematics during level walking," *J Orthop Res*, vol. 8, no. 3, pp. 383–392, 1990.

[106] C. Reinschmidt *et al.*, "Effect of skin movement on the analysis of skeletal knee joint motion during running," *J Biomech*, vol. 30, no. 7, pp. 729–732, 1997.

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

[108] M. Esch *et al.*, "Knee varus–valgus motion during gait – a measure of joint stability in patients with osteoarthritis?" *Osteoarthr Cartilage*, vol. 16, no. 4, pp. 522–525, 2008.

[109] A. Chang *et al.*, "Thrust during ambulation and the progression of knee osteoarthritis," *Arthritis Rheum*, vol. 50, no. 12, pp. 3897–3903, 2004.

[110] E. S. Grood and W. J. Suntay, "A Joint Coordinate System for the Clinical Description of {Three-Dimensional} Motions: Application to the Knee," *J Biomech Eng*, vol. 105, no. 2, pp. 136–144, 1983.

[111] J. Favre *et al.*, "A new ambulatory system for comparative evaluation of the three-dimensional knee kinematics, applied to anterior cruciate ligament injuries." *Knee Surg Sports Traumatol Arthrosc*, vol. 14, no. 7, pp. 592–604, 2006.

[112] A. Sabatini, "{Quaternion-Based} Extended Kalman Filter for Determining Orientation by Inertial and Magnetic Sensing," {*IEEE*} *Trans. Biomed. Eng.*, vol. 53, no. 7, pp. 1346–1356, 2006.

[113] S. Piazza and P. Cavanagh, "Measurement of the screw-home motion of the knee is sensitive to errors in axis alignment," *J Biomech*, vol. 33, no. 8, pp. 1029–1034, 2000.

[114] P. W. Stratford, D. M. Kennedy, and S. F. Robarts, "Modelling knee range of motion post arthroplasty: clinical applications." *Physiother Can*, vol. 62, no. 4, pp. 378–387, 2010.

[115] R. A. Bloomfield *et al.*, "Proposal and Validation of a Knee Measurement System for Patients With Osteoarthritis," *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 2, pp. 319–326, 2019.

[116] E. Yuksel *et al.*, "Assessing Minimal Detectable Changes and Test-Retest Reliability of the Timed Up and Go Test and the 2-Minute Walk Test in Patients With Total Knee Arthroplasty," *Journal of Arthroplasty*, vol. 32, no. 2, pp. 426–430, 2017.

[117] H. Akoglu, "User's guide to correlation coefficients," pp. 91–93, 2018.

[118] T. Conrozier *et al.*, "Getting Better or Getting Well? The Patient Acceptable Symptom State (PASS) Better Predicts Patient's Satisfaction than the Decrease of Pain, in Knee Osteoarthritis Subjects Treated with Viscosupplementation," *Cartilage*, vol. 9, no. 4, pp. 370–377, 2018.

[119] F. Tubach *et al.*, "Evaluation of clinically relevant states in patient reported outcomes in knee and hip osteoarthritis: The patient acceptable symptom state," *Annals of the Rheumatic Diseases*, vol. 64, no. 1, pp. 34–37, 2005.

[120] D. M. Kennedy *et al.*, "Preoperative Function and Gender Predict Pattern of Functional Recovery After Hip and Knee Arthroplasty," *Journal of Arthroplasty*, vol. 21, no. 4, pp. 559–566, 2006.

[121] N. F, SooHoo *et al.*, "Responsiveness of Patient Reported Outcome Measures in Total Joint Arthroplasty Patients," *The Journal of Arthroplasty*, vol. 30, no. 2, pp. 176–191, 2015.

[122] W. C. Lee *et al.*, "The minimal clinically important difference for Knee Society Clinical Rating System after total knee arthroplasty for primary osteoarthritis," *Knee Surgery, Sports Traumatology, Arthroscopy*, vol. 25, no. 11, pp. 3354–3359, 2017.

[123] C.-W. Ha *et al.*, "Increased Range of Motion Is Important for Functional Outcome and Satisfaction After Total Knee Arthroplasty in Asian Patients," *The Journal of Arthroplasty*, vol. 31, no. 6, pp. 1199–1203, 2016.

[124] D. J. Snyder *et al.*, "Preoperative Patient Reported Outcomes and Clinical Characteristics as Predictors of 90-day Cost/Utilization and Complications," *The Journal of Arthroplasty*, 2019.

[125] A. Siddiqi *et al.*, "Effect of Bundled Payments and Health Care Reform as Alternative Payment Models in Total Joint Arthroplasty: A Clinical Review," *The Journal of Arthroplasty*, vol. 32, no. 8, pp. 2590–2597, 2017.

[126] K. L. Urish *et al.*, "Predictors and Cost of Readmission in Total Knee Arthroplasty," *The Journal of Arthroplasty*, vol. 33, no. 9, pp. 2759–2763, 2018.

[127] J. M. Saucedo *et al.*, "Understanding Readmission After Primary Total Hip and Knee Arthroplasty: Who's at Risk?" *The Journal of Arthroplasty*, vol. 29, no. 2, pp. 256–260, 2014.

[128] Y.-H. Pua and P.-H. Ong, "Association of Early Ambulation With Length of Stay and Costs in Total Knee Arthroplasty: Retrospective Cohort Study," *American Journal of Physical Medicine and Rehabilitation*, vol. 93, no. 11, 2014.

[129] O. Babatunde *et al.*, "A Genetic Algorithm-Based Feature Selection," *International Journal of Electronics Communication and Computer Engineering*, vol. 5, no. 4, pp. 899–905, 2014.

[130] J. N. Mandrekar, "Receiver operating characteristic curve in diagnostic test assessment," *Journal of Thoracic Oncology*, vol. 5, no. 9, pp. 1315–1316, 2010.

[131] A. S. Desai *et al.*, "Manipulation for stiffness following total knee arthroplasty: when and how often to do it?" *European Journal of Orthopaedic Surgery and Traumatology*, vol. 24, no. 7, pp. 1291–1295, 2014.

[132] M. L. Barnett *et al.*, "Two-Year Evaluation of Mandatory Bundled Payments for Joint Replacement," *New England Journal of Medicine*, vol. 380, no. 3, pp. 252–262, 2019.

[133] S. M. Navarro *et al.*, "Machine Learning and Primary Total Knee Arthroplasty: Patient Forecasting for a Patient-Specific Payment Model," *Journal of Arthroplasty*, vol. 33, no. 12, pp. 3617–3623, 2018.

[134] R. K. Begg, M. Palaniswami, and B. Owen, "Support vector machines for automated gait classification," *IEEE Transactions on Biomedical Engineering*, vol. 52, no. 5, pp. 828–838, 2005.

[135] M. Huber, C. Kurz, and R. Leidl, "Predicting patient-reported outcomes following hip and knee replacement surgery using supervised machine learning," *BMC Medical Informatics and Decision Making*, vol. 19, no. 1, 2019.

[136] R. A. Bloomfield *et al.*, "Machine Learning Groups Patients by Early Functional Improvement Likelihood Based on Wearable Sensor Instrumented Preoperative Timed-Up-and-Go Tests," *Journal of Arthroplasty*, 2019.

[137] R. A. Bloomfield *et al.*, "Machine learning and wearable sensors at preoperative assessments: Functional recovery prediction to set realistic expectations for knee replacements," *Medical Engineering and Physics*, vol. 89, pp. 14–21, 2021.

[138] S. Patel *et al.*, "Monitoring motor fluctuations in patients with parkinsons disease using wearable sensors," *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 6, pp. 864–873, 2009.

[139] W. Rawat and Z. Wang, "Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review," *Neural Computation*, vol. 29, pp. 1–98, 2017.

[140] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *3rd International Conference on Learning Representations*, 2014.

[141] R. A. Bloomfield, M. G. Teeter, and K. A. McIsaac, "A Convolutional Neural Network Approach to Classifying Activities Using Knee Instrumented Wearable Sensors," *IEEE Sensors Journal*, vol. 20, no. 24, pp. 14 975–14 983, 2020.

Appendix A

Permissions and Approvals

This appendix contains forms and permission statements concerning the following:

- Permission statement from Elsevier for reuse of cited figures in Chapter 2.

- Permission statement from IEEE for some of the content of Chapter 4.

- Permission statement from Elsevier for some of the content of Chapter 6.

- Permission statement from Elsevier for some of the content of Chapter 6.

- Permission statement from IEEE for some of the content of Chapter 7.

- Ethics approvals for the human activity recognition study in Chapter 7 involving human subjects from the Human Subject Research Ethics Board at Western University.

- Information forms provided to subjects when personally recruiting subjects for participation in the Chapter 7 and obtaining informed consent.

- Ethics approval for the functional instrumentation of patients in the clinic using wearable sensors for data collected by Megan Fennema in Chapter 5.

- Ethics approval for the functional instrumentation of patients in the clinic using wearable sensors for data collected by Jordan Broberg in Chapter 6.

- Ethics approval for the functional instrumentation of patients in the clinic using wearable sensors for data collected by Harley Williams in Chapter 6.

RightsLink®

Home   ? Help   ✉ Email Support   👤 Riley Bloomfield ⌄

Modification of the Grood and Suntay Joint Coordinate System equations for knee joint flexion

Author: Danè Dabirrahmani, Michael Hogg
Publication: Medical Engineering & Physics
Publisher: Elsevier
Date: January 2017

Order Completed

Thank you for your order.

This Agreement between Mr. Riley Bloomfield ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

Your confirmation email will contain your order number for future reference.

License Number	5030400366104
License date	Mar 15, 2021

🖶 Printable Details

✅ Licensed Content

Licensed Content Publisher	Elsevier
Licensed Content Publication	Medical Engineering & Physics
Licensed Content Title	Modification of the Grood and Suntay Joint Coordinate System equations for knee joint flexion
Licensed Content Author	Danè Dabirrahmani, Michael Hogg
Licensed Content Date	Jan 1, 2017
Licensed Content Volume	39
Licensed Content Issue	n/a
Licensed Content Pages	4
Journal Type	HS

📖 Order Details

Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Format	electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No

📄 About Your Work

Title	Graduate Student
Institution name	Western University
Expected presentation date	Mar 2021

📁 Additional Data

Portions	Fig. 1

📍 Requestor Location

Requestor Location

📇 Tax Details

Publisher Tax ID	GB 494 6272 12

$ Price

Total	0.00 CAD

Total: 0.00 CAD

CLOSE WINDOW   ORDER MORE

Copyright Clearance Center

RightsLink®

Home   Help   Email Support   Sign in   Create Account

Proposal and Validation of a Knee Measurement System for Patients With Osteoarthritis

Author: Riley Aaron Bloomfield

Publication: Biomedical Engineering, IEEE Transactions on

Publisher: IEEE

Date: Feb. 2019

Copyright © 2019, IEEE

Thesis / Dissertation Reuse

The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you may print out this statement to be used as a permission grant:

Requirements to be followed when using any portion (e.g., figure, graph, table, or textual material) of an IEEE copyrighted paper in a thesis:

1) In the case of textual material (e.g., using short quotes or referring to the work within these papers) users must give full credit to the original source (author, paper, publication) followed by the IEEE copyright line © 2011 IEEE.
2) In the case of illustrations or tabular material, we require that the copyright line © [Year of original publication] IEEE appear prominently with each reprinted figure and/or table.
3) If a substantial portion of the original paper is to be used, and if you are not the senior author, also obtain the senior author's approval.

Requirements to be followed when using an entire IEEE copyrighted paper in a thesis:

1) The following IEEE copyright/ credit notice should be placed prominently in the references: © [year of original publication] IEEE. Reprinted, with permission, from [author names, paper title, IEEE publication title, and month/year of publication]
2) Only the accepted version of an IEEE copyrighted paper can be used when posting the paper or your thesis on-line.
3) In placing the thesis on the author's university website, please display the following message in a prominent place on the website: In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of [university/educational entity's name goes here]'s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.

If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation.

BACK   CLOSE WINDOW

RightsLink®

Machine Learning Groups Patients by Early Functional Improvement Likelihood Based on Wearable Sensor Instrumented Preoperative Timed-Up-and-Go Tests

Author:
Riley A. Bloomfield,Harley A. Williams,Jordan S. Broberg,Brent A. Lanting,Kenneth A. McIsaac,Matthew G. Teeter

Publication: The Journal of Arthroplasty

Publisher: Elsevier

Date: October 2019

© 2019 The Author(s). Published by Elsevier Inc.

BACK   CLOSE WINDOW

RightsLink®

Machine learning and wearable sensors at preoperative assessments: Functional recovery prediction to set realistic expectations for knee replacements

Author: Riley A. Bloomfield,Jordan S. Broberg,Harley A. Williams,Brent A. Lanting,Kenneth A. McIsaac,Matthew G. Teeter

Publication: Medical Engineering & Physics

Publisher: Elsevier

Date: March 2021

Journal Author Rights

BACK

CLOSE WINDOW

RightsLink®

IEEE
Requesting permission to reuse content from an IEEE publication

A Convolutional Neural Network Approach to Classifying Activities Using Knee Instrumented Wearable Sensors

Author: Riley A. Bloomfield

Publication: IEEE Sensors Journal

Publisher: IEEE

Date: 15 Dec.15, 2020

Copyright © 2020, IEEE

Thesis / Dissertation Reuse

The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you may print out this statement to be used as a permission grant:

Requirements to be followed when using any portion (e.g., figure, graph, table, or textual material) of an IEEE copyrighted paper in a thesis:

1) In the case of textual material (e.g., using short quotes or referring to the work within these papers) users must give full credit to the original source (author, paper, publication) followed by the IEEE copyright line © 2011 IEEE.
2) In the case of illustrations or tabular material, we require that the copyright line © [Year of original publication] IEEE appear prominently with each reprinted figure and/or table.
3) If a substantial portion of the original paper is to be used, and if you are not the senior author, also obtain the senior author's approval.

Requirements to be followed when using an entire IEEE copyrighted paper in a thesis:

1) The following IEEE copyright/ credit notice should be placed prominently in the references: © [year of original publication] IEEE. Reprinted, with permission, from [author names, paper title, IEEE publication title, and month/year of publication]
2) Only the accepted version of an IEEE copyrighted paper can be used when posting the paper or your thesis on-line.
3) In placing the thesis on the author's university website, please display the following message in a prominent place on the website: In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of [university/educational entity's name goes here]'s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.

If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation.

BACK   CLOSE WINDOW

Activity recognition based on lower extremity orientations

Principal Investigator:
Dr. Matthew Teeter

This letter of information describes the research study and your role as the participant. The purpose of this letter is to provide you with information required for you to make an informed decision regarding participation in this research. Please read this form carefully. Do not hesitate to ask anything about the information provided.

Study Purpose

Wearable sensors have become more available, smaller in size, and efficient for analyzing movement. Their components are located in most personal electronics including smart watches and phones.

Measurement of the legs can be used to evaluate hip and knee function and can enable researchers and clinicians to monitor levels and types of activities before or after surgical procedures such as hip or knee replacement. The amount and types of activities performed before surgery may give insight into expected outcomes after surgery. You are being invited to test these sensors as a healthy volunteer. By using wearable sensors, data can be recorded in many environments which will make data collection easier for future studies.

The purpose of this study is to determine if the orientations of the legs measured from wearable sensors can be used to correctly classify daily activities to accurately estimate activity levels.

Procedure

Twenty healthy volunteers will be recruited to participate in this study. It is expected to take approximately 30 minutes of your time. If you agree to participate you will meet at the Western University Thompson Recreation and Athletic Centre and be asked to complete the following activities:

1. Walking for 30 seconds at both a normal and slower than normal pace on a treadmill
2. Walking for 30 seconds over-ground
3. Running or jogging for 45 seconds at both a normal and slower than normal pace on a treadmill
4. Standing vertically for 45 seconds facing two different directions
5. Sitting in a chair for 45 seconds facing two different directions
6. Sitting cross-legged for 45 seconds facing two different directions
7. Ascending a stair climbing machine for 45 seconds at both a normal and slower than normal pace, placing each arm on their respective handrails for stability
8. Descending a set of stairs for 90 seconds using a single arm on a handrail for support
9. Biking on a stationary bike for 45 seconds at both a normal and slower than normal pace
10. Deep squatting for 30 seconds with knees bent as much as possible and only feet touching the ground
11. Kneeling on both knees for 30 seconds with back straight and upright (perpendicular to floor)
12. Kneeling on both knees for 30 seconds with forehead near the floor and arms stretched in front
13. Kneeling on a single knee (simulating tying a shoe) for 45 seconds on each knee.
14. Laying on back, left side, right side for 30 seconds each

There will be no restrictions on the pace of these tests so long as they are executed at two different speeds. You will be able to practice completing any of the tasks prior to undergoing any of the testing.

While completing these activities, you will be instrumented with a set of four lightweight wearable sensors. There will be one sensors mounted above and below each knee, attached with stretchable hook and latch straps. A picture of the sensor with its strap can be seen below as well as a with a quarter for size reference:

An observer will attach the wearable sensors using elastic hook and latch straps above and below each of your knees. The observer will also monitor and record sensor data using an iPod touch during activity completion.

Risks
There is a risk of falls with this study, however, we expect this risk to be minimal as only healthy volunteers will participate in this study and all participants will be given the opportunity to practice executing the required activities prior to undergoing any of the testing.

Benefits
Participation in this study will not provide any known benefit to you.

Compensation
There will be no compensation provided for your participation in this study. Parking costs associated with participating in this study will be reimbursed.

Voluntary Participation
Your participation in this study is voluntary. You may leave the study at any time. If you decide to withdraw from the study, the information that was collected before you left the study will still be used in order to help answer the research question. No new information will be collected without your permission.

Alternatives to Study Participation
An alternative to the study procedures described above is to not participate in this study.

Confidentiality
No identifiable information will be collected as part of this research study. In any publication, presentation, or report, any information that would reveal your identify will not be published. Study data will be kept for a minimum of 7 years.

You will be given a copy of this letter of information and consent form once it has been signed. You do not waive any legal rights by signing the consent form. Representatives of the University of Western Ontario Health Sciences Research Ethics Board may access the data to monitor the conduct of the research.

Western

Activity recognition based on lower extremity orientations

Informed Consent

Agreement of Participating Subject
I have read the accompanying letter of information, have had the nature of the study explained to me and I agree to participate. All questions have been answered to my satisfaction.

_____

Print participant's full name

_____

Participant's signature   Date

_____

Name of person obtaining consent

_____

Signature of person obtaining consent   Date

Date: 5 July 2018

To: Matthew Teeter

Project ID: 112091

Study Title: Activity recognition based on lower extremity orientations

Application Type: HSREB Initial Application

Review Type: Delegated

Full Board Reporting Date: 17July2018

Date Approval Issued: 05/Jul/2018 11:47

REB Approval Expiry Date: 05/Jul/2019

_____

Dear Matthew Teeter

The Western University Health Science Research Ethics Board (HSREB) has reviewed and approved the above mentioned study as described in the WREM application form, as of the HSREB Initial Approval Date noted above. This research study is to be conducted by the investigator noted above. All other required institutional approvals must also be obtained prior to the conduct of the study.

Documents Approved:

Document Name	Document Type	Document Date	Document Version
activityRecognitionInformationConsent	Written Consent/Assent	03/Jul/2018	1,2
activityRecognitionStudyProtocol	Protocol		
dataCollectionFile	Other Data Collection Instruments	30/May/2018	1.0
par-q	Paper Survey	03/Jul/2018	1.0
phoneScript	Telephone Script	28/Jun/2018	1.0

No deviations from, or changes to, the protocol or WREM application should be initiated without prior written approval of an appropriate amendment from Western HSREB , except when necessary to eliminate immediate hazard(s) to study participants or when the change(s) involves only administrative or logistical aspects of the trial.

REB members involved in the research project do not participate in the review, discussion or decision.

The Western University HSREB operates in compliance with, and is constituted in accordance with, the requirements of the TriCouncil Policy Statement: Ethical Conduct for Research Involving Humans (TCPS 2); the International Conference on Harmonisation Good Clinical Practice Consolidated Guideline (ICH GCP); Part C, Division 5 of the Food and Drug Regulations; Part 4 of the Natural Health Products Regulations; Part 3 of the Medical Devices Regulations and the provisions of the Ontario Personal Health Information Protection Act (PHIPA 2004) and its applicable regulations. The HSREB is registered with the U.S. Department of Health & Human Services under the IRB registration number IRB 00000940.

Please do not hesitate to contact us if you have any questions.

Sincerely,

Nicola Geoghegan-Morphet, Ethics Officer on behalf of Dr. Joseph Gilbert, HSREB Chair

Note: This correspondence includes an electronic signature (validation and approval via an online system that is compliant with all regulations).

Western
Research
Western University Health Science Research Ethics Board

HSREB Delegated Initial Approval Notice

Principal Investigator: Dr. Brent Lanting
Department & Institution: Schulich School of Medicine and Dentistry\Orthopaedic Surgery, London Health Sciences Centre

Review Type: Delegated
HSREB File Number: 109398
Study Title: The relationship between patient reported outcome measures and variability in wearable sensor-implemented Timed-Up-and-Go tests after total knee arthroplasty

HSREB Initial Approval Date: June 29, 2017
HSREB Expiry Date: June 29, 2018

Documents Approved and/or Received for Information:

Document Name	Comments	Version Date
Revised Western University Protocol	Received June 22, 2017	
Letter of Information & Consent		2017/05/26
Instruments	UCLA Activity Score	2017/05/29
Data Collection Form/Case Report Form	Received May 30, 2017	

The Western University Health Science Research Ethics Board (HSREB) has reviewed and approved the above named study, as of the HSREB Initial Approval Date noted above.

HSREB approval for this study remains valid until the HSREB Expiry Date noted above, conditional to timely submission and acceptance of HSREB Continuing Ethics Review.

The Western University HSREB operates in compliance with the Tri-Council Policy Statement Ethical Conduct for Research Involving Humans (TCPS2), the International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use Guideline for Good Clinical Practice Practices (ICH E6 R1), the Ontario Personal Health Information Protection Act (PHIPA, 2004), Part 4 of the Natural Health Product Regulations, Health Canada Medical Device Regulations and Part C, Division 5, of the Food and Drug Regulations of Health Canada.

Members of the HSREB who are named as Investigators in research studies do not participate in discussions related to, nor vote on such studies when they are presented to the REB.

The HSREB is registered with the U.S. Department of Health & Human Services under the IRB registration number IRB 00000940.

_____
Ethics   half of Dr. Marcelo Kremenchutzky, HSREB Vice Chair
EO: Erika Basile ___ Grace Kelly ___ Katelyn Harris___ Nicola Morphet ___ Karen Gopaul ___ Patricia Sargeant ___

Western
Research Western University Health Science Research Ethics Board
HSREB Full Board Initial Approval Notice

Principal Investigator: Dr. Douglas Naudie
Department & Institution: Schulich School of Medicine and Dentistry\Surgery,London Health Sciences Centre

Review Type: Full Board
HSREB File Number: 109512
Study Title: Radiostereometric analysis of gap balancing versus measured resection for the Journey II total knee replacement

HSREB Initial Approval Date: October 02, 2017
HSREB Expiry Date: October 02, 2018

Documents Approved and/or Received for Information:

Document Name	Comments	Version Date
Revised Western University Protocol	Received July 25, 2017	
Revised Letter of Information & Consent		2017/07/25
Data Collection Form/Case Report Form	RSA Data Collection Form (29-Jun-2017)	
Data Collection Form/Case Report Form	KSS (Clinician) Questionnaire	
Data Collection Form/Case Report Form	KSS (Patient) Questionnaire	
Instruments	WOMAC Questionnaire	2017/06/22
Instruments	UCLA Activity Score Questionnaire	2017/06/22
Instruments	Short Form-12 (SF-12) Questionnaire	2017/06/22
Recruitment Items	Recruitment Letter	
Other	Registration NCT# 03290170	

The Western University Health Science Research Ethics Board (HSREB) has reviewed and approved the above named study, as of the HSREB Initial Approval Date noted above.

HSREB approval for this study remains valid until the HSREB Expiry Date noted above, conditional to timely submission and acceptance of HSREB Continuing Ethics Review.

The Western University HSREB operates in compliance with the Tri-Council Policy Statement Ethical Conduct for Research Involving Humans (TCPS2), the International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use Guideline for Good Clinical Practice Practices (ICH E6 R1), the Ontario Personal Health Information Protection Act (PHIPA, 2004), Part 4 of the Natural Health Product Regulations, Health Canada Medical Device Regulations and Part C, Division 5, of the Food and Drug Regulations of Health Canada.

Members of the HSREB who are named as Investigators in research studies do not participate in discussions related to, nor vote on such studies when they are presented to the REB.

The HSREB is registered with the U.S. Department of Health & Human Services under the IRB registration number IRB 00000940.

_____
Ethics Officer, on behalf of Dr. Joseph Gilbert, HSREB Chair
EO: Erika Basile____ Grace Kelly ___ Katelyn Harris___ Nicola Morphet ___ Karen Gopaul ___ Patricia Sargeant ___

LAWSON FINAL APPROVAL NOTICE

LAWSON APPROVAL NUMBER: R-17-327

PROJECT TITLE: The Impact of Total Knee Arthroplasty Surgical Technique with Cementless Implants on Coronal Plane Motion

PRINCIPAL INVESTIGATOR: Dr. Brent Lanting

LAWSON APPROVAL DATE: Tuesday, 5 September 2017

Health Sciences REB#:   109486

ReDA ID: 2921

Please be advised the above project was reviewed by Lawson Administration and the project:

Was Approved

Please provide your Lawson Approval Number (R#) to the appropriate contact(s) in supporting departments (eg. Lab Services, Diagnostic Imaging, etc.) to inform them that your study is starting. The Lawson Approval Number must be provided each time services are requested.

Dr. David Hill
V.P. Research
Lawson Health Research Institute

All future correspondence concerning this study should include the Lawson Approval Number and should be directed to Sherry Paiva, Research Approval Officer, Lawson Health Research Institute, 750 Baseline Road, East, Suite 300.

cc: Administration

Curriculum Vitae

Name

Riley Aaron Bloomfield

Education

PhD Electrical and Computer Engineering, Biomedical Systems   2016 - 2021
University of Western Ontario, London, CA

BESc Software Engineering, with distinction   2012 - 2016
University of Western Ontario, London, CA

Publications

Bloomfield, R., Broberg, J., Williams, H., Lanting, B., McIsaac, K., Teeter, M. "Machine Learning and Wearable Sensors at Preoperative Assessments: Functional Recovery Prediction to Set Realistic Expectations for Knee Replacements," *Medical Engineering & Physics,* vol. 89, pp. 14-21, 2021.

Polus, J., Bloomfield, R., Vasarhelyi, E., Lanting, B., Teeter, M. "Machine Learning Predicts the Fall Risk of Total Hip Arthroplasty Patients Based on Wearable Sensor Instrumented Performance Tests," *Journal of Arthroplasty,* vol. 36, no. 2, pp. 573-578, 2021.

Bloomfield, R., Teeter, M., McIsaac, K., "A Convolutional Neural Network Approach to Classifying Activities Using Knee Instrumented Wearable Sensors," *IEEE Sensors Journal,* vol. 20, no. 24, pp. 14975-14983, 2020.

Van de Kleut, M., Bloomfield, R., Teeter, M., Athwal, G. "Monitoring daily shoulder activity before and after reverse total shoulder arthroplasty using inertial measurement units," *Journal of Shoulder and Elbow Surgery,* 2020

Bloomfield, R., Williams, H., Broberg, J., Lanting, B., McIsaac, K., Teeter, M., "Machine Learning Groups Patients by Early Functional Improvement Likelihood Based on Wearable Sensor Instrumented Preoperative Timed-Up-and-Go Tests," *Journal of Arthroplasty,* vol. 34,

no. 10, pp. 2267-2271, 2019.

Bloomfield, R., Fennema, M., McIsaac, K., Teeter, M., "Proposal and Validation of a Knee Measurement System for Patients with Osteoarthritis," *IEEE Transactions on Biomedical Engineering,* vol. 66, no. 2, pp. 319-326, 2019.

Fennema, M., Bloomfield, R., Lanting, B., Birmingham, T., Teeter, M., "Repeatability of measuring knee flexion angles with wearable inertial sensors," *Knee,* vol. 26, no. 1, pp. 97-105, 2019.

Awards

Ion Inculet PhD Award in Engineering   Apr 2020
Department of Electrical & Computer Engineering, Western University
Awarded annually to a single graduate student registered in a PhD program, based on academic achievement and research effort.

Top 10 Research Advances of 2019   Nov 2019
Arthritis Society
In recognition of the publication: Bloomfield et al., "Machine learning groups patients by early functional improvement likelihood based on wearable sensor instrumented preoperative timed-up-and-go tests"

Queen Elizabeth II Graduate Scholarship in Science and Technology   May 2019
Province of Ontario & Western University
Awarded but declined in favour of NSERC Postgraduate Scholarship

Graduate Student Research Award   May 2019
Department of Electrical & Computer Engineering, Western University

Outstanding Presentation in Graduate Symposium (Biomedical Systems)   May 2019
Department of Electrical & Computer Engineering, Western University

NSERC Postgraduate Scholarship - Doctoral   Apr 2019
Natural Sciences and Engineering Research Council of Canada

Ion Inculet PhD Award in Engineering   Mar 2019
Department of Electrical & Computer Engineering, Western University

Outstanding Presentation in Graduate Symposium (Biomedical Systems))   May 2018
Department of Electrical & Computer Engineering, Western University

Ontario Graduate Scholarship   Sep 2017
Province of Ontario & Western University

Top Undergraduate Software Engineering Final Project - Team   Apr 2016
Western University, London, ON
Awarded by a committee of department faculty and invited industry judges.

NSERC Undergraduate Student Research Award   May 2015
Natural Sciences and Engineering Research Council of Canada

Western Entrance Scholarship   Sep 2012
Western University, London, ON

Conferences

Podium Presentations

Bloomfield, R., Williams, H., Broberg, J., Lanting, B., Teeter, M., "Machine Learning Groups Total Knee Replacement Patients by Likelihood of Short-Term Functional Improvement Post-Surgery Based on Preoperative Instrumented Timed-Up-and-Go Tests," *International Society of Technology in Arthroplasty Annual Congress,* Oct 2019. Toronto, ON.

Bloomfield, R., McIsaac, K., Teeter, M., "Classification of Static and Dynamic Activities From a Leg-Instrumented Wearable Sensor System," *International Society of Technology in Arthroplasty Annual Congress,* Oct 2019. Toronto, ON.

Bloomfield, R., "Application of a shallow convolutional neural network for multivariate time series activity data classification," *Western Electrical & Computer Engineering Research Symposium,* May 2019. London, ON. **Awarded** *Best Presentation* in the Biomedical Systems category.

Bloomfield, R., "Portable Measurement and Analysis of Patients with Knee Osteoarthritis," *Western Electrical & Computer Engineering Research Symposium,* May 2018. London, ON. **Awarded** *Best Presentation* in the Biomedical Systems category.

Bloomfield, R., "Knee joint motion measurement during the timed up and go test using low-cost wearable sensors," *Imaging Network Ontario Symposium,* Mar 2017. London, ON.

Poster Presentations

Bloomfield, R., Broberg, J., Lanting, B., Teeter, M., "Predicting Functional Recovery to Realistically Set Patient Expectations Following Total Knee Replacement," *Orthopaedic Research Society Meeting,* Feb 2021, Delivered virtually due to to COVID-19.

Polus, J., Bloomfield, R., Vasarhelyi, E., Lanting, B., Teeter, M., "Machine Learning Predicts Fall Risk of Total Hip Arthroplasty Patients at Six-Weeks Post-Operation Based on Wearable Sensor Instrumented Performance Tests at Two-Week Clinical Appointments," *Canadian Orthopaedic Association Meeting,* Jun 2020, Halifax, NS. Postponed due to COVID-19.

Bloomfield, R., Broberg, J., Williams, H., Lanting, B., Teeter, M., "Machine Learning of Wearable Sensor Instrumented Preoperative Timed-up-and-go Tests Predicts Functional Recovery Likelihood Following Total Knee Replacement," *American Association of Orthopaedic Surgeons Annual Meeting,* Mar 2020. Orlando, FL. Displayed virtually due to COVID-19.

Bloomfield, R., Williams, H., Broberg, J., Lanting, B., Teeter, M., "Machine learning categorizes total knee replacement patients by likelihood of functional improvement at three-months post-surgery based on preoperative instrumented timed-up-and-go tests," *International Conference on Ambulatory Monitoring of Physical Activity and Movement,* Jun 2019. Maastricht, Netherlands.

Bloomfield, R., McIsaac, K., Teeter, M., "Application of a Shallow Convolutional Neural Network for Activity Recognition from Leg-Instrumented Wearable Sensors," *International Conference on Ambulatory Monitoring of Physical Activity and Movement,* Jun 2019. Maastricht, Netherlands.

Williams, H., Bloomfield, R., Broberg, J., Howard, J., Vasarhelyi, E., Naudie, D., Lanting, B., Teeter, M., "Wearable Sensors: Providing a New Perspective to Post-Operative TKA Functional Recovery," *International Combined Orthopaedic Research Societies Annual Meeting,* Jun 2019.

Bloomfield, R., Williams, H., Lanting, Teeter, M., "Machine learning for early recovery prediction after total knee replacement based on preoperative instrumented timed-up-and-go tests," *London Health Research Day,* Apr 2019. London, ON.

Bloomfield, R., Williams, H., Lanting, Teeter, M., "Machine learning functional metrics for recovery path analysis following total knee replacement," *Imaging Network Ontario Symposium,* Mar 2019. London, ON. **Awarded** *Honorable Mention* in scientific poster competition.

Williams, H., Bloomfield, R., Yuan, X., Howard, J., Lanting, B., Teeter, M., "Cementless Total Knee Arthroplasty: Does surgical technique impact implant migration?," *Imaging Network Ontario Symposium,* Mar 2019. London, ON.

Teeter, M., Bloomfield, R., Fennema, M., Webster, J., Williams, H., Naudie, D., Howard, J., and Lanting, B. "Wearable sensors for evaluating total knee arthroplasty patient performance during the timed up and go functional test," *Canadian Orthopaedic Research Society Meeting,* Jun 2018. Victoria, BC.

Perelgut, M., Bloomfield, R., Vasarhelyi, E., Lanting, B., Teeter, M., "Total Hip Arthroplasty Surgical Approach and Implant Design: Effects on Patient Function and Implant Migration,"

Robarts Research Retreat, Jun 2018. London, ON.

Williams, H., Bloomfield, R., Lanting, B., Teeter, M., "Wearable Sensors: Identifying new temporal and joint-specific metrics to improve the characterization of patient functional capacity pre- and post-TKA," *Western Department of Surgery Research Day,* Jun 2018. London, ON.

Bloomfield, R., Fennema, M., Lanting, B., McIsaac, K., Teeter, M., "Gait asymmetry detection using wearable sensors for osteoarthritis patients," *Canadian Bone & Joint Conference,* May 2018. London, ON.

Perelgut, M., Bloomfield, R., Vasarhelyi, E., Lanting, B., Teeter, M., "Determining the effectiveness of using wearable sensors to measure the functional status of total hip arthroplasty patients," *Canadian Bone & Joint Conference,* May 2018. London, ON.

Williams, H., Bloomfield, R., Yuan, X., Howard, J., Lanting, B., Teeter, M., "Wearable Sensors: Identifying Temporal and Quantitative Joint-Specific Metrics to Improve the Assessment of Patient Functional Status Pre- and Post- TKA," *Canadian Bone & Joint Conference,* May 2018. London, ON.

Bloomfield, R., Fennema, M., Lanting, B., McIsaac, K., Teeter, M., "Quantifying knee range of motion using wearable sensors in osteoarthritis patients," *London Health Research Day,* May 2018. London, ON.

Bloomfield, R., Fennema, M., Lanting, B., McIsaac, K., Teeter, M., "Gait asymmetries observed during segmented walking activities in patients with knee osteoarthritis," *Imaging Network Ontario Symposium,* Mar 2018. Toronto, ON.