

# Leveraging Ubiquitous Sensing for Quantifying the Quality of Motion in Mobile Health

**Robert Gutierrez**

## **Committee Members:**

Professor Mehdi Boukhechba (Advisor)

Professor Afsaneh Doryab (Chair)

Professor Laura Barnes

Professor Joe Hart

Doctor Rebecca Scharf

A dissertation presented for the degree of

Doctor of Philosophy

Engineering Systems and Environment

University of Virginia

July 2022

# APPROVAL SHEET

This  
Dissertation  
is submitted in partial fulfillment of the requirements  
for the degree of  
Doctor of Philosophy

Author: Robert Gutierrez

This Dissertation has been read and approved by the examining committee:

Advisor: Mehdi Boukhechba

Advisor:

Committee Member: Afsaneh Doryab

Committee Member: Laura Barnes

Committee Member: Joe Hart

Committee Member: Rebecca Scharf

Committee Member:

Committee Member:

Accepted for the School of Engineering and Applied Science:



Jennifer L. West, School of Engineering and Applied Science

August 2022

---

To *mom*, for she taught me to be never give up

To *dad*, for he taught me there's always a solution

To *my kids*, for they kept me motivated to make them proud

To *my wife*, for I am reminded of how lucky I am to have her love and support

The views expressed in this dissertation are those of the author and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the U.S. Government.

---

## ACKNOWLEDGMENTS

I would like to express my gratitude to the committee members for their time and support. First, I want to thank Professor Boukhechba and Professor Barnes for their support over the last few years. Thank you for taking a chance on me. It has been my pleasure to work with and learn from you both! I would also like to thank my committee members, Professor Doryab, Professor Hart, and Dr. Scharf, for their guidance and discussions in the projects I had the honor to be a part of. Their expertise and passion in their respective fields encouraged me to continually strive for excellence. I would also like to thank Shashwat Kumar for his collaboration, support, and insights on multiple projects.

Most importantly, I want to thank my family for their continuous love and support. My parents for always pushing me to keep going. My kids for their smiles that served as my inspiration. And my wife, for her endless support, motivation, and countless sacrifices.

---

## ABSTRACT

Ubiquitous sensing from smartphones and wearable devices have proven to be useful for a variety of applications, from sports to modern medicine. These devices are embedded with powerful sensors which track the motion of an object in real time. From this data, quantitative measures of motion can be extracted while also assessing the quality of that movement.

Motor function assessments are designed to evaluate human performance, but the outcome measures of these assessments are often based on subjective or ordinal rating scales that do not keep track of motion quantitatively. In this dissertation, we leverage ubiquitous sensing and signal-based curve matching to quantify the quality of motion through the development of an objective scoring measure . We then validate our measure with a novel study involving the functional assessment of patients with neuromuscular disorders.

We also demonstrate the feasibility of ubiquitous sensing combined with signal feature analysis to detect and assess rehabilitation progress in patients post-ACL reconstruction surgery. Finally, we present a visualization framework that combines analytical modeling with comparative trajectory analysis to model motion data into meaningful information related to the quantity and quality of motion.

# Table of Contents

<b>List of Tables</b>	<b>vi</b>
<b>List of Figures</b>	<b>vii</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Literature Review</b>	<b>5</b>
2.1 Summary . . . . .	5
2.2 Ubiquitous Sensing . . . . .	5
2.3 Quantitative Motion Modeling using Time Series Analysis . . . . .	9
2.3.1 Signal Feature Analysis . . . . .	9
2.3.2 Curve Matching . . . . .	13
2.4 Motion Quality Assessments . . . . .	15
2.5 Current Gaps and Limitations . . . . .	19
<b>3 A Framework for Modeling Deviations in Motion Trajectories</b>	<b>23</b>
3.1 Background and Motivation . . . . .	23
3.2 Methodology . . . . .	24
3.2.1 Feature Representation of Motion Data . . . . .	25
3.2.2 Motion Alignment with Curve Matching . . . . .	25
3.2.3 Time Window and Statistical Feature Selection . . . . .	25
3.2.4 Trajectory Divergence Analysis . . . . .	26
3.2.5 Visualization . . . . .	26

---

3.3	Discussion . . . . .	29
<b>4</b>	<b>Validation through Human Motion Capture with Wearable Sensors</b>	<b>30</b>
4.1	Leveraging Wearable Sensors to Detect and Assess Rehabilitation Progress . . . . .	30
4.1.1	Background and Motivation . . . . .	30
4.1.2	Study Design . . . . .	32
4.1.3	Methodology . . . . .	34
4.1.4	Results . . . . .	37
4.1.5	Applied MISTA Framework . . . . .	50
4.1.6	Discussion . . . . .	54
4.1.7	Limitations . . . . .	55
4.1.8	Conclusion . . . . .	55
4.2	A Study on Quantifying Motion Quality with Wearable Sensors . . . . .	55
4.2.1	Background and Motivation . . . . .	56
4.2.2	Methodology . . . . .	58
4.2.3	Results . . . . .	65
4.2.4	MISTA Framework Application . . . . .	77
4.2.5	Future Work . . . . .	78
4.2.6	Conclusion . . . . .	80
<b>5</b>	<b>Discussion</b>	<b>82</b>
5.1	Clinical Relevance . . . . .	83
5.2	Limitations . . . . .	84
5.3	Future Works . . . . .	85
<b>6</b>	<b>Conclusion</b>	<b>87</b>
	<b>List of Publications</b>	<b>88</b>

# List of Tables

4.1	<b>Extracted features</b> . . . . .	37
4.2	Iso 180 Mann-Whitney-U Statistical Tests . . . . .	41
4.3	Iso 90 Mann-Whitney-U Statistical Tests . . . . .	44
4.4	<b>Classification Accuracy for Iso 180</b> . . . . .	46
4.5	<b>Classification Accuracy for Iso 90</b> . . . . .	46
4.6	<b>Classification Accuracy for Combined Exercises</b> . . . . .	47
4.7	Cohort Profile . . . . .	62
4.8	Traditional Motor Function Assessment Scores . . . . .	63
4.9	List of activities performed by each participant. . . . .	64
4.10	Extracted features . . . . .	66
4.11	DTW-MeND vs ED scores for 3 angular velocities. . . . .	66
4.12	Classification Accuracy for Arm Curl . . . . .	75
4.13	Classification Accuracy for Arm Rotation . . . . .	76

# List of Figures

2.1	Illustration of sensor-based HAR [1] . . . . .	6
2.2	Example of a current motor function assessment form . . . . .	20
2.3	Example of a clinical setting for motor function assessment . . . . .	21
2.4	Vicon MX <sup>®</sup> Motion Capture System. . . . .	22
3.1	<b>MISTA Framework.</b> Our novel visualization framework can models deviations in motion trajectory. . . . .	24
3.2	<b>Divergence Analysis.</b> The dot size can be changed depending on the user’s choice for comparing the statistical feature. The top plot (a) shows the dot size relative to the absolute difference between the time windows of the template and query trajectories. The bottom plot (b) illustrates larger dots for the time windows that are deemed statistically significant after applying a statistical test to compare the time windows. Here, we show the use of a one-way ANOVA test with $\alpha = 0.05$ to increase the dot size by 3x. . . . .	27
3.3	<b>MISTA Gait Example.</b> As an example, we apply MISTA to a healthy participant’s gait cycle as they walk on the treadmill. . . . .	28
4.1	<b>Sensor placement.</b> Delsys Trigno sensors were applied bilaterally at a measured distance on the vastus lateralis for tracking gait kinematics and quadriceps muscle activations. . . . .	34

---

4.2	<b>Methodology breakdown.</b> Our methodology begins with data collection, followed by muscle activation detection, and feature extraction. Then, we provide a descriptive analysis of the motions, apply predictive modeling to classify the motions into their respective cohorts, and then apply our MISTA framework for trajectory analysis of limb symmetry. . . . .	35
4.3	<b>Iso 180 EMG data</b> The muscle activations for healthy subject 1 during the iso 180 motion are shown with the involved leg on the left and the uninvolved leg on the right. . . . .	35
4.4	<b>Muscle activation separation.</b> The muscle activations of healthy subject 1’s involved leg during the iso 180 exercise are separated into individual partitions for comparative analysis. . . . .	36
4.5	<b>Wavelength (WL) feature boxplot for iso 180.</b> For the iso 180 exercise, the median WL values of the involved and uninvolved legs within the healthy cohort are nearly identical while the median WL value for the patients’ uninvolved leg is shown to be much higher than their involved leg. . . . .	38
4.6	<b>Iso 180 feature comparison.</b> . . . . .	40
4.7	<b>Zero-Crossing (ZC) feature boxplot for iso 90.</b> Here we see that within the healthy cohort, the median ZC value is very similar between the involved and uninvolved leg. However, there appears to be a clear difference between the median ZC values for the patient cohort. . . . .	42
4.8	<b>Iso 90 feature comparison.</b> . . . . .	43
4.9	<b>Classification confusion matrices.</b> Confusion matrices for the top classification models in the healthy versus patient scenario. The left is the random forest model for the iso 180 exercise while the right is the random forest model for the iso 90 exercise. . . . .	45

---

4.10	<b>KDE plots for variables provided.</b> The top two plots are the distributions from the IKDC And KOOS self-reported assessments while the bottom plots are the distributions from the maximal volitional isometric contraction (MVIC) flexion and extension metrics provided by the Biodex isokinetic dynamometer. . . . .	47
4.11	<b>KDE plots of iso 180 features.</b> Here we see that within the healthy cohort, the median ZC value is very similar between the involved and uninvolved leg. However, there appears to be a clear difference between the median ZC values for the patient cohort. . . . .	48
4.12	<b>KDE plots of iso 90 features.</b> Here we see that within the healthy cohort, the median ZC value is very similar between the involved and uninvolved leg. However, there appears to be a clear difference between the median ZC values for the patient cohort. . . . .	49
4.13	Comparison between Patient 1’s involved and uninvolved leg. . . . .	50
4.14	Comparison between the average patient involved leg and the average healthy leg. . . . .	51
4.15	Comparison between Patient 1’s involved leg and the average healthy leg. Here we also include all the healthy gait cycles in the background. . . . .	52
4.16	Comparison of multiple gait cycles between Patient 1’s uninvolved leg and healthy participant 1’s dominant leg. . . . .	52
4.17	The sEMG data during the same gait cycle for each participant . . . . .	53
4.18	Visual representation of generating MeND scores . . . . .	59
4.19	The Brooke Upper Extremity Scale is a 6-point scale used to classify upper extremity [2]. Higher scores indicate more severe motion impairment. This scale was used as ground truth for all participants. . . . .	63

---

4.20	Pilot study set up with two MbiEntLab MMR+ sensors placed on a participant's hands and the Microsoft Kinect v2 on a tripod to record and track movements. A custom built sensory board was designed to test specific upper extremity motions related to activities of daily living (ADLs). . . . .	64
4.21	Arm curl motion performed at 3 different angular velocities (fast/normal/slow).	67
4.22	The top boxplot (a) displays the DTW-MeND curl scores for each cohort. With a p-value $< 0.001$ , there is a significant difference between the control cohort and the DMD cohort, as well as between the control cohort and the SMA cohort. However, there was no significant difference found between the DMD and SMA cohorts with a p-value = 0.334. The bottom boxplot (b) depicts the ED curl scores for each cohort. With a p-value $< 0.001$ , there is a significant difference between the control cohort and the DMD cohort. There was no significant difference found between the DMD and SMA cohorts with a p-value = 0.088, nor was there a significant difference between the control and SMA cohorts with a p-value = 0.095. . . . .	68
4.23	The top boxplot (a) displays the DTW-MeND rotation scores for each cohort. A significant difference between the Control and DMD cohorts was determined with a p-value = 0.020, as well as a significant difference between the Control and SMA cohorts with a p-value = 0.037. However, no significant difference was found between DMD and SMA cohorts, with a p-value=0.077. The bottom boxplot (b) depicts the ED rotation scores for each cohort. With p-values $> 0.05$ , there were no significant difference found between any of the cohorts. . . . .	69
4.24	Sample comparison of DTW-MeND (a) and ED scores (b) for Arm Curl. The vertical lines represent the cohort average values for the specific activity and scoring measure. . . . .	71

---

4.25	Sample comparison of DTW-MeND (a) and ED scores (b) for Arm Rotation. The vertical lines represent the cohort average values for the specific activity and scoring measure. . . . .	72
4.26	DTW-MeND quantifies the quality of motion subjectively, to indicate how similar the motion is to the control cohort. From top to bottom, the patients' normalized distance score of the arm curl increases as the quality of the curl motion decreases. . . . .	73
4.27	Confusion matrices for the arm curl Logistic Regression models. DTW-MeND on the left and ED on the right. . . . .	75
4.28	Confusion matrices for the arm rotation Logistic Regression models. DTW-MeND on the left and ED on the right. . . . .	76
4.29	Raw arm curl motions performed by Patient 2, with blue as their left arm and orange as their right arm. The black line is the average healthy arm curl. . .	78
4.30	The MISTA framework applied to Patient 2's arm curl motions, with the left arm on top and the right arm on bottom. . . . .	79

# Chapter 1

## Introduction

Advancements in ubiquitous sensing have redefined the landscape for human activity recognition (HAR) by enabling the emergence of personal, group, and community scale sensing applications. These technologies are already revolutionizing many sectors of our economy, including business, healthcare, social networks, environmental monitoring, and transportation. Specifically, smartphones and wearable devices with embedded inertial measurement units (IMUs) can be leveraged with machine learning techniques to detect and assess behavioral performance, as noted by [3]. Living in a data-driven age, these commercially available devices provide valuable information into a person's health and wellness, that can be used for recreational activities and medical applications [4, 5, 6, 7, 8].

According to CCS Insight, the demand for smart wearable devices is expected to grow 24% worldwide from 2020 to 2021, with an estimated 1.2 billion devices by 2025 [9]. Wearable IMUs offer many advantages in their inherent design (compact, comfortable, inexpensive, waterproof, etc) and are able to track users in a natural environment. IMUs are embedded with a suite of sensors, that may include a triaxial accelerometer, a triaxial gyroscope, surface electromyography (sEMG), and a variety of other sensors [7]. Motion capture with these devices are becoming essential for research areas focused on sports performance analysis, rehabilitation, and various medical applications. One of the leading companies for motion capture analysis, Vicon Motion Systems ©, maintains that IMU-based motion analysis is

---

driving human motion optimization, both in the field and in the lab. Motion analysis is more than just measuring one's movement, but also visualizing and analyzing the data in a manner that provides feedback for understanding differences, such as left versus right or current status to a baseline. The best motion tracking systems rely on a combination of devices (wearables, cameras, etc.) which are costly and impractical for capturing everyday motion. Quantitative feedback from IMUs can inform coaching, assist in monitoring training performance, and help evaluate the quality of a movement [10].

Quality of motion is defined as the manner in which a motion is performed, usually in comparison to an established reference. Then, measuring the quality of motion provides the means to evaluate the effectiveness of interventions (or lack thereof) and monitor progress in more detail [11]. For healthcare applications, clinical researchers will be able to measure the efficacy of investigative therapies, adjust focus and investments to promising developments, and mitigate future investments on non-performing studies, thereby speeding discovery. Then, the development of a more objective scoring measure could be deployed for continuous, home-based monitoring, through the use of wearable ubiquitous solutions.

Motor function assessments provide a quantitative baseline for evaluating human performance, which can aid in understanding the severity of impairment and monitoring recovery during rehabilitation. Most motion assessments have been developed to evaluate a user performing a specific set of motions, where the outcome measures of these assessments are based on subjective or ordinal rating scales that do not keep track of quantitative measures of motion quality. This causes the outcome measures to be susceptible to a ceiling effect [12]. Clinicians and researchers alike recognize that the reliance on the human eye to facilitate and score a user's motion is not the most effective means for detecting small changes in patient progress, which often go unnoticed [13]. For example, for patients with neuromuscular disorders, these assessments are generally performed in the clinic, with the motions bearing no resemblance to activities of daily living (ADLs) [14, 15]. Some of these assessments are no longer relevant, as treatments have enabled patients to live longer and be able to gain

---

function [15]. Additionally, the current metrics are not sensitive to smaller changes, where fine motor skills of the hands and fingers may be the remaining ability [15]. These limitations also extend to traditional rehabilitation assessments, where a patient may be asked to perform a motion they are unable to complete, and may not receive credit the motion they could accomplish. Yet, these limitations can be mitigated by quantifying the quality of motion.

There has also been little emphasis on developing new visualizations for IMU motion trajectory. Erbaugh [16] noted that visual feedback plays a pivotal role in one’s ability to make correct movements. In the case of range of motion (ROM) assessment, visualization empowers both the doctor and patient to be more informed on one’s motion that would otherwise be lost in a single quantitative number. Therefore, by visualizing deviations in motion trajectory, quantitative measures of motion can be interpreted while also assessing the quality of that movement.

This dissertation is focused on leveraging ubiquitous sensing to quantify the quality of motion through the development of an objective scoring measure and develop a graphical approach to visualize deviations along a motion trajectory. We accomplish this by addressing the following research questions:

**Research Question 1:** Is it feasible to leverage ubiquitous sensing to quantify the quality of motion?

**Research Question 2:** Which features can be derived from ubiquitous sensors to characterize the quality of motion?

**Research Question 3:** How to model deviations in motion trajectory patterns?

The remainder of this work is presented as follows. In chapter 2, we summarize related works on motion quality assessments, ubiquitous sensing, and quantitative motion modeling. In chapter 3, we first present our novel scoring measure to quantify the quality of motion that

---

builds upon the foundation of curve matching. To validate our measure, we introduce our study on using wearable sensors to track upper extremity motor function in patients with neuromuscular disorders. In chapter 4, we present our work on leveraging wearable sensors to detect and assess rehabilitation progress for patients recovering from ACL reconstruction surgery. In chapter 5, we present our framework for modeling deviations in motion trajectory. Finally, in chapter 7, we discuss the clinical relevance of our work, the limitations, future research directions, and conclusion.

## Chapter 2

# Literature Review

### 2.1 Summary

To understand the limitations of quantifying motion with wearable sensors, we first begin with a review of existing ubiquitous modalities in section 2.2. Section 2.3 describes two overarching methods for analyzing motion data, *Signal Feature Analysis* and *Curve Matching*. We then discuss motion quality assessments in section 2.4, focusing on those relevant to the applications we will be covering in this thesis.

### 2.2 Ubiquitous Sensing

Human activity recognition (HAR) with ubiquitous sensors is a growing field revolutionizing many sectors of the world, including business, healthcare, social networks, environmental monitoring, and transportation. Specifically, mobile sensing systems such as smartphones and wearable inertial measurement units (IMUs) have proven to be an effective means for motion tracking applications. Embedded with a suite of powerful sensors, such as accelerometer, gyroscope, and magnetometer, these devices leverage pervasive computing to provide valuable information into a person's health and wellness, that can be used for recreational activities and medical applications [4, 5, 6, 7, 8].

Accelerometers measure the linear acceleration in a specific direction and are primarily used to estimate velocity and position through integration. However, the standard ac-

celerometer reading includes the force of gravity which Gyroscopes measure angular velocity to describe the rotational movement of a device while magnetometers measure the force of Earth’s magnetic field on the device. These sensors are designed to measure the static and dynamic states of object. Static state parameters include spatial position, orientation, angles between parts, etc. while dynamic state parameters consist of displacement, trajectory, velocity, acceleration, etc [17]

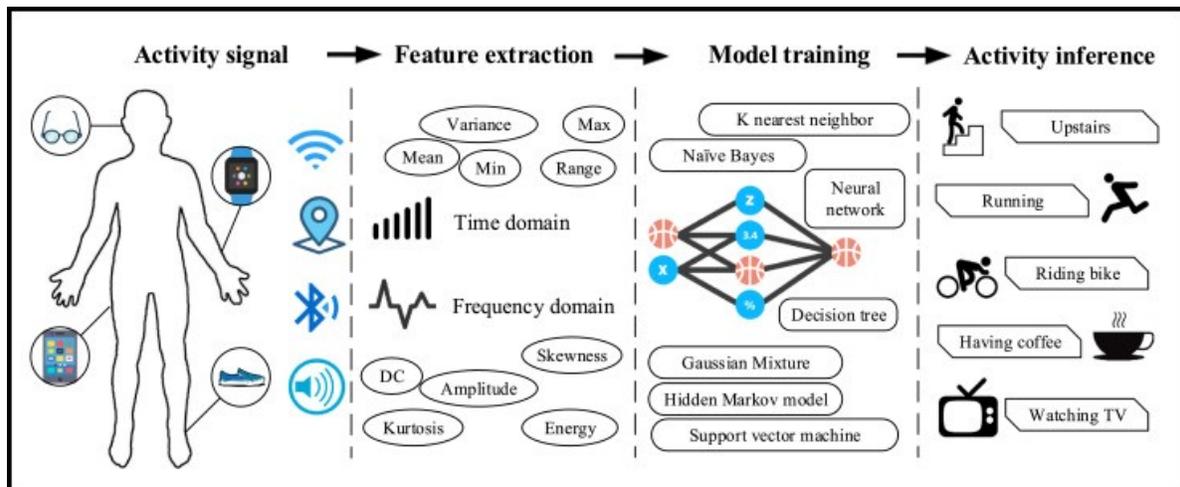


Fig 2.1. Illustration of sensor-based HAR [1]

The illustration shown in in 2.1 shows the process of how modern technology classifies human motions, by collecting data from a multitude of sensors, extracting representative features, training a model, and inferring the associated activities. Most research on motion analysis with IMUs either uses multiple devices (combination of wearables, cameras, and edge devices) or focuses on a set of gestures for classification [18, 19, 20, 21].

As wearable technology becomes more ubiquitous, mHealth applications have grown immensely. Wearable IMUs sensors are often embedded in smartphones, smartwatches, and many application specific devices. The number of commercially available IMU devices is growing exponentially and often include other sensors of interest such as GPS, EMG, ECG, heart rate, etc. The wrist-worn device space is extremely saturated with more companies developing smartwatches and fitness trackers. Surface Electromyography (sEMG) devices are another ubiquitous device gaining more traction as they have been shown to recognize

---

hand and arm gestures while worn on the forearm by recording electrical impulses stemming from muscular contractions through the use of surface electrodes [22]. Often times, sEMGs are paired with accelerometers or IMUs to better capture motion. The list of IMU embedded devices continues to grow, as more are created for specific applications, such as gait biomechanics and swimming performance analysis.

The advantages of using a wearable IMU include mobility, ease of use, reduced cost, and ubiquitous nature. However, IMUs are not without their own limitations where an accelerometer accumulates error over time or "drifts" while a magnetometer may suffer from magnetic interference. IMUs are also limited in their ability to recognize and track fine-grained movements with one's hands and fingers. Camera based systems are considered the "gold standard" for HAR but issues of privacy, cost and lack of mobility make have limited their widespread adoption outside a research space.

Ubiquitous sensing also includes the use of camera based technologies. Vision-based activity recognition systems, such as the Microsoft Kinect and Leap Motion systems, use a camera, IR sensor, and IR depth sensor to extract features that can be used for object detection, motion analysis, and other characteristics. While many motion tracking cameras exist in the commercial sector, the vast majority of research for skeletal motion tracking leverages the Microsoft Kinect (v1 and v2) [23, 6, 24, 25, 26]. The "gold standard" for camera based motion analysis is utilizing optical motion tracking through technologies found in the Vicon MX<sup>®</sup> motion capture system that uses 20 cameras to track a series of reflective markers placed all over an individual's body. The main limitations with systems such as these are their cost, confinement to a specific area, and privacy, making them impractical for capturing everyday motion.

The authors in [27] also recognized the lack of attention on the development of metrics for performance evaluation. Vakanski et al [28] proposed a series metrics for use in physical therapy, based on using a reference model as ground truth for comparison. Su et al [29] leveraged the Microsoft Kinect system with Dynamic Time Warping (DTW) and fuzzy logic

---

to build a fuzzy inference model that includes a physician's subjective evaluation of a patient's performance for home-based rehabilitation. Similarly, Yu and Xiong [30] use the Kinect along with Xsens wearable sensors to capture exergame motion. Their work mapped bone vectors in 3D space and applied DTW to yield a quantitative score. However, they also developed a performance score by converting the DTW distance to a percentage based on the number of their problem set.

For sports, there has been plenty of research focused on gait analysis [8], with favorable results for the increase in performance due to wearable IMU feedback. However, limited work has focused directly on upper extremity movements. "Advancements in wearables have allowed individual athletes, sports teams, and physicians to monitor functional movements, workloads, and biometric markers to maximize performance. Coaches can utilize this data to personalize training and conditioning to match the athlete's current physical status based on trends on their data. [31]"

Medical applications leveraging IMUs for motion tracking have become quite prevalent over the last decade. One medical application for 3D motion analysis includes measuring the quantity and quality of movement for extremity rehabilitation and motor function assessments. In the case of neuromuscular diseases, many patients suffer from reduced upper extremity function. Current motor function assessments do not capture all positive progress and only concerned on whether a patient can or cannot complete a specific task. This technology has the ability to measure ROM while also capturing micro-movements that have previously gone overlooked. Similarly, ROM assessments extend to clinical orthopaedics as studied by [32], where an IMU device can aid in goniometer evaluations by providing accurate measurements and eliminating the inter-rater reliability issue. Additionally, this technology allows a patient to perform these motor function assessments in the comfort of their own home, rather than in a clinical environment. As [33] stated, wearable IMUs allow "for real-time movement tracking in real-world settings." Especially when providing care to young children, a clinical setting may affect their demeanor, which may inaccurately reflect their

---

progress on an assessment.

Telemedicine, the practice of evaluating, diagnosing, and treating patients remotely, is becoming more of a viable option due to the proliferation of motion capture devices. The benefits of telemedicine include less travel time, lower costs, and more comfort for the patient. One form of telerehabilitation is through exergames, exercise combined with games) that leverages wearable IMUs to track user movement of personalized exercises for a specified treatment plan [3]. One example, the MIRA (Medical Interactive Recovery Assistant) rehabilitation platform, [34] leverages the Kinect sensor to capture user’s movements of therapeutic exercises, tracking the trajectories of the joints. However, there are limitations with the current platforms not able to track micro-movements or specific motions. However, these exergames would benefit from the addition of IMU sensors that would provide more accurate position and orientation information for each movement.

Pairing quantitative outcomes with metrics on the quality of the movements can provide informative data for rehabilitation interventions and monitoring recovery. The following section will provide a more in depth review on the quantification of motion.

## **2.3 Quantitative Motion Modeling using Time Series Analysis**

### **2.3.1 Signal Feature Analysis**

Time series signal feature analysis for quantitative motion modeling can be done through two main approaches: *knowledge-driven* and *data-driven*.

#### **2.3.1.1 Knowledge-driven**

Knowledge-driven techniques are generally rule-based or case-based to produce a model that structurally represents the context of the data. This approach is based on the notion of leveraging expert knowledge about the motion to detect known behavior, or the expected ranges the motion will occur. The expert would extract specific features for classification related to an attribute. For example, one could compute the standard deviation of speed as a measure of motion smoothness and then visually identify the trajectories based on a set of

---

rules or conditions that have been learned from experience. In addition, the expert may be able to provide situational context by incorporating information not available to the sensor. The advantages of using knowledge-driven techniques include interpretability, low-false positive rate, and are context-aware [35]. However, they are limited in their ability to adapt to new data or anomalies and require human effort to maintain.

**Feature Extraction** Feature extraction is a technique used by both approaches, varying in the manner the features are derived. These features are quantitative measures, usually based on a time window [4]. With a knowledge-driven approach, the features are selected from experience and domain knowledge, causing them to be data specific. For gait motion analysis, typical features include stride length, stride duration, gait velocity, step counts, etc. Then for upper extremity motor function analysis, standard features include the mean, variance, peak velocity, range of motion (RoM), pairwise correlation, spectral energy, and mean absolute deviation.

The set of features can then be compared directly, yielding a vector of difference values between two trajectories. However, the problem with using these features are that they do not necessarily capture the quality of the motion. Quality of motion describes the movement smoothness and whether the movement was performed correctly. Motion quality features may include movement efficacy, shape tracing, log dimensionless jerk (LDLJ), spectral arc length (SPARC), range of motion (RoM), etc. Despite several attempts to define quantitative measure of movement smoothness, there is no consensus on the best technologies nor the most appropriate measures to use for a specific motion [36].

Traditional time series analysis focuses on signal classification techniques, which first extract a set of features or transform the signal into a new representation. Then, a distance metric is used to measure the similarity between two signals. The extracted features, whether they are in the time or frequency domain, can either be representative of the entire trajectory or they can be based on sets of points that span the trajectory. Features derived from the time domain include mean absolute value (MAV), zero crossing rate, autoregressive

---

model coefficients while features in the frequency domain include mean frequency (MNF) and median frequency (MDF) [18, 37]. Other common features include signal norm, signal energy, signal variance, frequency analysis, zero crossing rate, entropy, etc. The set of features can then be compared directly, yielding a vector of difference values between two trajectories. However, the problem with using these features are that they do not necessarily capture the quality of the motion. Quality of motion describes the movement smoothness and whether the movement was performed correctly. Motion quality features may include movement efficacy, shape tracing, log dimensionless jerk (LDLJ), spectral arc length (SPARC), range of motion (RoM), etc. Despite several attempts to define quantitative measure of movement smoothness, there is no consensus on the best technologies nor the most appropriate measures to use for a specific motion [36].

### **2.3.1.2 Data-driven**

Feature extraction from data-driven techniques derive features and patterns from the data directly through machine learning (ML) and data mining [38]. ML approaches can generally be categorized as supervised, unsupervised, and semi-supervised learning. Unlike the knowledge-driven approach, this technique can be generalized to new data, allowing the analysis to be replicated across multiple conditions and patients. However, data-driven approaches tend to have more false positives, may be uninterpretable, and can require large amounts of data for training the model. To understand what data-driven methods should be considered for time series analysis, we will first make clear the type of data that we are working with. In our case, the data is discrete, equally spaced, and stochastic. Generally, this type of time series data can be analysed through the following techniques: discrete point analysis, feature extraction, transformation or representation, and curve matching.

The two approaches do not need to be mutually exclusive, as the integration of knowledge-driven models with data-driven or machine learning models can overcome the limitations to produce an informed, dynamic model [38].

While many features can be extracted, the number of features need to be considered as

---

high dimensionality increases run time and complexity. Dimensionality reduction techniques such as factor analysis and Principal Components Analysis (PCA) can be used to find the optimal features, while t-stochastic neighbor embedding (t-SNE) plots can aid in visualizing the high dimensional features for clusters. The PCA algorithm is able to process all the extracted features and project them into vectors relative to their variance. These new vectors can then be treated as features for classification.

Classification techniques can then be used for analyzing sensor data. Some common classifiers include decision trees (DT), k-nearest neighbors (kNN), random forests (RF), support vector machine (SVM), bootstrap aggregation classification and regression trees (Bagged CART), boosted logistic regression (BLG), gradient boosting machines (GBM), and neural networks (NN). One of the main strengths for deep learning through neural networks is in “its ability to automatically extract features in task dependent manner” instead of the reliance on hand-crafted features [39]. A drawback for deep learning stems from its “black box” approach, where interpretability of the model is lost.

**Discrete Point Analysis** Discrete point analysis relies on finding specific points on a curve and comparing their magnitude, such as the max, min, inflection points, etc. This type of analysis is good for instances when one already knows that those specific points are of interest. However, this method ignores a vast majority of the data and may not provide enough information if condensed to single number.

**Time Series Representation** Wearable IMUs have the potential to fill the gap where video based motion tracking falters. Many attempts have been made to use a IMU devices for tracking motion but very few focus on visualizing the trajectory of that motion. Through information visualization, quantitative measures of motion can be evaluated while also assessing the quality of that movement. In the case of range of motion assessment, visualization empowers both the doctor and patient to be more informed on one’s motion that would otherwise be lost in a single quantitative number. Erbaugh [16] noted that visual feedback plays a pivotal role in one’s ability to make correct movements.

---

### 2.3.2 Curve Matching

Given two trajectories, two trajectories can be compared directly through the use of a distance metric to produce a single value that represents the similarity between them. However, there must be careful consideration when selecting a distance metric. The chosen distance metric should have the capability to handle noisy sample points and trajectory shifts, such as time stretch shifts and scale suppresses. By appropriately handling these circumstances, a distance metric can accurately distinguish between two trajectories along a path with similar shapes [29]. Su et al conducted a survey of trajectory distance measures, evaluating 15 distance measures based on data type (discrete versus continuous) and whether the measure considers temporal information [29].

One similarity metric that can be used is the cross-correlation as defined by Equation 2.1. This equation, commonly recognized as the Pearson product-moment correlation coefficient, measures the temporal similarity of two curves [40].

$$\rho_{xy}(\ell) = \frac{\sum_{i=0}^{N-1} (x_i - \bar{x}) * (y_{i-\ell} - \bar{y})}{\sqrt{\sum_{i=0}^{N-1} (x_i - \bar{x})^2} \sqrt{\sum_{i=0}^{N-1} (y_{i-\ell} - \bar{y})^2}} \quad (2.1)$$

This metric can be good for detecting phase shifts between two curves, but requires them to be of equal length.

If looking at the sequences directly or spatially, one can use distance measures like ED, Dynamic Time Warping (DTW) or derivative DTW. These measures find sample point pairs along two trajectories by minimizing the transformation cost. Then, the cumulative cost yields the final distance. The matching process can either be complete matches where every point is matched, including duplicates, or partial matches which means some points are not paired at all. Within the complete matching metrics, the most common are ED and DTW. ED is defined as the summation of the ordered point pair distances, shown below:

---


$$d_{ED} = \sqrt{\sum_{k=1}^P (x_{ik} - v_{jk})^2} \quad (2.2)$$

However, ED has several limitations, especially in its ability to handle signal transformations, such as shifting and scaling, and the requirement for the sequences to be the same length. Möller-Levet et al. [41] introduced a novel similarity measure, short time series distance (STS), that leverages temporal information from time-series data by treating each time-series as piece-wise linear functions and calculating the difference of the slopes. Mathematically, it is defined as:

$$d_{STS} = \sqrt{\sum_{k=1}^P \left( \frac{v_{j(k+1)} - v_{jk}}{t_{(k+1)} - t_k} - \frac{x_{i(k+1)} - x_{ik}}{t_{(k+1)} - t_k} \right)^2} \quad (2.3)$$

Similar to ED, STS is sensitive to scaling but can be overcome with a z-score standardization of the time-series. Since equal length trajectories are a rare occurrence, the length limitation can be overcome through either the use of sliding windows or more flexible measurements. DTW overcomes the length limitation by calculating a warping path that minimize the distance between pairs of indices, subject to a few restrictions [42]. These restrictions include monotonicity, continuity, and boundary conditions (constrained endpoints).

$$dtw(i, j) = c(x_i, y_j) + \min \begin{cases} dtw(i-1, j) \\ dtw(i, j-1) \\ dtw(i-1, j-1) \end{cases} \quad (2.4)$$

Equation 2.4 [43] details the recurrence relation of the DTW method, which allows it to be robust to shifts in the data. Thus, sum of the distance between the indices is reported as the similarity distance. The two main limitations with DTW are its computational complexity and alignment singularities. One solution to overcoming the singularities issue is to take the derivative [44]. Rather than using ED as the distance measure between two points, derivative DTW (DDTW) takes the square of the distance of the derivatives for the two

---

matched points.

Time Alignment Measurement (TAM), proposed by Folgado et al. [45], is a similarity measurement designed to "characterize temporal misalignment" between two signals, by incorporating temporal information rather than signal amplitude. Their work first introduces an adjustment to DTW with the inclusion of sliding windows (SW-DTW) to account for singularities by aligning distinct features, such as local minima, maxima or valleys. By changing the cost function from element-to-element distance to a window distance, the number of singularities are reduced and a better alignment is achieved. Then, TAM computes the cost to warp the optimal path by characterizing the windows when "the series are in phase, advance or delay" [45]. The formal definition for this measurement is:

$$\Gamma = \overrightarrow{\psi} + \overleftarrow{\psi} + (1 - \overline{\psi}), \quad \Gamma \in \{\mathbb{R}_0^+ | \Gamma \in [0 : 3]\} \quad (2.5)$$

where  $\overrightarrow{\psi}$  is the fraction of advance,  $\overleftarrow{\psi}$  is the delay, and  $\overline{\psi}$  is the phase. TAM can be calculated from the result of any signal alignment technique, given that each element of the series has a pairwise relation. Despite the focus on TAM, the author's use of SW-DTW may prove to be more useful for the alignment of the two signals.

Wang et al. proposed an IMU-based gait normalcy index (INI) that leveraged gait variables and kinematics to measure the similarity between a patient and the normal mean of healthy persons [46]. Their INI was based off a PCA method developed by Schutte et al. [47] that is sensitive to both the deviation of each gait variable and the overall relationship between the variables. By leveraging PCA, the method is able to remove the correlation between variables that would otherwise occur if distance values were directly assessed from the original variables.

## 2.4 Motion Quality Assessments

The assessment of motion quality has garnered significant interest as it can provide coaches, trainers, and clinicians with valuable information regarding an individual's ability to perform

---

a specific movement. Motor function assessments can provide a quantitative baseline for evaluating the performance and are essential to monitoring recovery during rehabilitation. This has led to the development of several assessments tools, many of which are designed for a specific population or objective. These tools often use simple Likert-scale scoring systems (i.e. 0-5) which limits their sensitivity to small changes in movement quantity and quality [14]. To quantify motion within a medical setting, clinicians have developed several motor function assessments for specific use cases. For example, upper extremity impairment commonly affects patients with varying medical conditions, including cerebral palsy, spinal muscular atrophy (SMA), Duchenne muscular dystrophy (DMD), Parkinson's, and hemiplegia, as well those recovering from surgery as in the case of breast cancer survivors and spinal-cord related injuries. With the varying degree of impairment for each of those conditions mentioned, researchers and clinicians have developed assessments for each condition. Understandably, the motions that matter most will differ across some of the diagnoses, hence the need for separate assessments, but the motions will also vary from patient-to-patient. The patient-to-patient variation here does not refer to the individualized preference of motion for increased quality of life, but rather the difference in severity of a patient's underlying diagnosis. For example, the range of motor function impairment for children with SMA include those who can move independently and those whose only movement is a small flexion of the wrist.

Some of the most common tools include the Children's Hospital of Philadelphia Infant Test of Neuromuscular Disorders (CHOP-INTEND) [48], Hammersmith Functional Motor Scale Expanded (HFMSE) [49], the Brooke upper extremity scale [50], Fugl-Meyer Assessment (FMA) [51], Wolf Motor Function Test (WMFT) [52], Shriner's Hospital for Children Upper Extremity Evaluation (SHUEE) [53], Upper Limb Module (ULM), and many more Likert-type scales.. Some assessments are self administered surveys, such as the Disabilities of Arm, Shoulder, and Hand (DASH) questionnaire. Each assessment has its own set of motions to perform along with its own evaluation process, and clinicians often have differing opinions on which assessment is the "gold standard". These tests are conducted

---

and scored by human examiners, which introduces variability and uncertainty in evaluation scores. Traditional assessments can be tedious and time consuming, which is why previous and ongoing research have tried to implement ubiquitous technologies, such as smartphones, smartwatches, cameras and other wearable sensors, to aid in the evaluation process and provide a path toward home-based assessment evaluations [54, 55, 56, 57, 58, 59, 6, 60, 61].

The DASH [62] is a self report questionnaire consisting of 30 questions with the goal of measuring symptoms and functional status of upper extremity musculoskeletal conditions [62]. The questions vary between asking a patient if they are able to perform a variety of activities, such as opening a jar, turning a key, and push open a door, as well as subjective questions about whether they are in pain or have been limited in their daily activities. The responses for this questionnaire range from 1 to 5, where 1 coincides with "No Difficulty" while a 5 represents "Unable". The assessment generally lasts for 5-10 minutes and is intended for patients over 18 years old. While the DASH can be used for a quick determination of a patient having an impairment, it does not capture any movements, thus it is extremely limited in its use for rehabilitation recommendation and recovery progress.

The FMA [51] is an evaluation tool that assesses voluntary movement, reflex activity, grasp, and coordination. This particular assessment contains both upper and lower extremity portions. A few of the movements included within the upper extremity portion are shoulder abduction, gripping a tennis ball, elbow extension, and bicep reflex activity. Due to its large adaptation within the medical and research communities, it is frequently referred to as the "gold standard" in upper extremity motor function assessment. However, as [52] mention, the FMA can be difficult to use and has become outdated for new rehabilitation treatments. The time to complete this assessment ranges from 30-35 minutes. A clinician observes the performance of each movement and provides a score from 0 to 2, where a value of 0 indicates the patient could not perform the movement, a value of 1 if the movement was performed partially, and a value of 2 if the movement was performed fully. While these movements provide a broad evaluation of each upper extremity joint, there are no

---

functional tasks associated with the evaluation. Since the movements are not measured on a quantitative scale nor are they timed, the assessment provides little information for tracking improvements over time. The WMFT [52] is a tool that quantifies upper extremity movement through timed motions and functional tasks [58]. With a total of 17 tasks and a time limit of 120 seconds per task, each are listed in order of complexity to assess overall movements and the speed at which they are performed. Thus, two scores are recorded: 1) WMFT-Time, the time required to complete the tasks; and 2) WMFT-Functional Ability Scale (FAS), scores ranging from 0 to 5 based on a patient’s functional capacity to perform each task. On average, the test has a duration between 30-45 minutes and is performed on a single side. Some of the tasks include, placing one’s hand on a box, lifting a pencil, and turning a key. As a requirement, the WMFT must be video recorded as the FAS scores are determined post hoc by a trained clinician. The Upper Limb Module (ULM) that was developed as an international effort between clinicians, physical therapists, and researchers to address the shortcomings in previous SMA motor function assessments [63]. The ULM was targeted towards younger children and overcame the flooring effects found in the Hammersmith Functional Motor Scale-Expanded (HFMSSE). However, the ULM had issues with the ceiling effect which led to the Revised Upper Limb Module (RULM) [63]. A few of the tasks that were added in the RULM include picking up coins, pushing a button light with one hand, and raising a 200 gram cup to the mouth. While the tasks were related to ADLs and allowed for complete ROM with the different joints, the RLM does not explicitly measure each task.

Next, we turn to motion quality assessments for sports. Traditional sports performance analysis is done through the process of a coach or expert watching the every move of an athlete to assess their strengths, weaknesses, and technical form. With the vast amount of data available from motion analysis, an athlete’s performance can be optimized to determine how, when, and where a player should move. The difference of a second or an inch could have a substantial effect, which is why the interpretation of that data is so crucial. For

---

swing analysis in sports such as golf, baseball, and tennis, an athlete would wear an IMU device on their dominant wrist to gather data as they swing to hit the ball. Motion quality would assess characteristics of wrist position and orientation, acceleration profile, and various angles. Similarly, the applications for swim stroke motion assessment have expanded as smartwatches from Garmin, Apple, and Samsung and other IMU devices can measure various swimming metrics such as stroke count, velocity, and biomechanical performance [64]. For ACL related injuries, functional assessments revolve around leg isolated tests so that the results of each leg can be compared against one another, known as a limb symmetry index (LSI). Sports therapists use a threshold of 90% as part of the criteria for clearing a patient to return-to-sport (RSI). The isolated tests include muscle strength testing with an isokinetic dynamometer (Biodex Medical Systems Inc., NY, USA), hop tests, single leg squat, 6 meter hop for time, and others. Some functional assessment exercises assess both legs, like the box jump landing to vertical jump, but are reserved for patients who are much further along in their rehabilitation.

However, sports medicine specialists also rely on patient reported functional assessment surveys, like the International Knee Documentation Committee subjective form (IKDC) and Knee Injury and Osteoarthritis Outcome Score (KOOS). Both assessments have been found to be reliable measures for patients who have undergone ACLR surgery [65, 66, 67, 68]. The IKDC is based on 0-to-100 scale, with higher scores translating to more knee function while the KOOS is the combination of five subscales: pain, symptoms, activities of daily living, sport and recreation activities, and quality of life. Each subscale has questions scored between 0 and 5, which are then aggregated on a 0-to-100 scale, with higher scores indicating less knee issues.

## **2.5 Current Gaps and Limitations**

To grasp the need for our work presented in this thesis, this section will discuss the limitations of existing studies in relation to wearable sensors and motion quality assessments. Given the current technology available for wearable devices, the largest criticism is drawn from the



---

therapy, and may not translate to activities of daily living. Often times, these rooms are furnished with expensive equipment, like that shown in Figure 2.3. After speaking with parents and physicians, many note that the performance of the patient in the clinic is not always representative of their true functional ability. Therefore, a decrease in performance can be attributed to fatigue, nerves, or anxiety stemming from the environment the patient is testing in.



**Fig 2.3.** Example of a clinical setting for motor function assessment

The third main limitation is complexity. The "gold standard" for motion analysis is optical motion tracking through technologies found in the Vicon MX<sup>®</sup> motion capture system that uses 20 cameras to track a series of reflective markers placed all over an individual's body. An example of this is shown in Figure 2.4. Systems such as these are complex, expensive, confined to a specific area, and are impractical for capturing everyday motion.



**Fig 2.4.** Vicon MX<sup>®</sup> Motion Capture System.

## Chapter 3

# A Framework for Modeling Deviations in Motion Trajectories

### 3.1 Background and Motivation

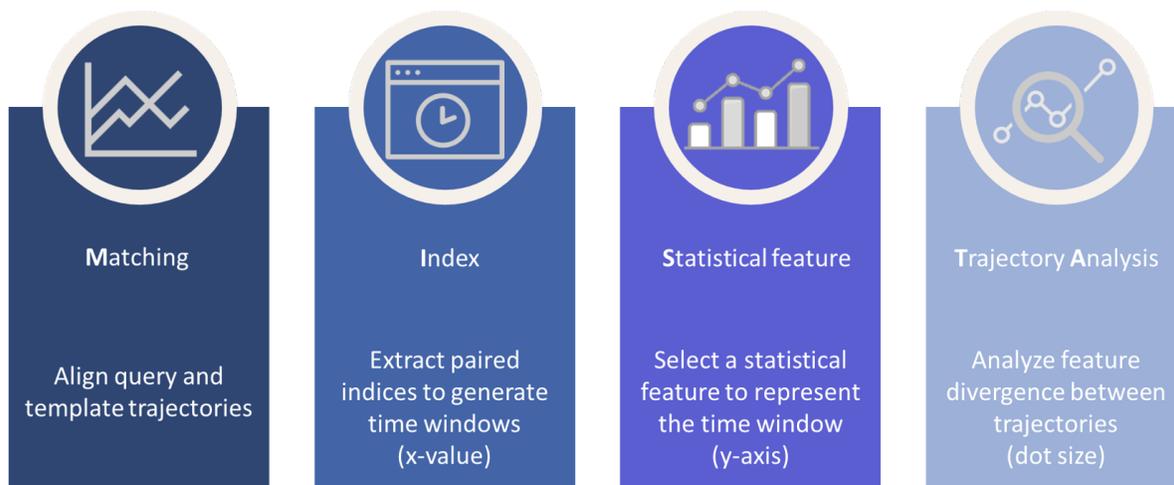
Modeling motion trajectory is important in a variety of applications such as sports performance analysis, gaming, animation, and healthcare. Though cost and the adoption of new technologies are barriers to entry for most healthcare applications, low-cost wearable inertial measurement units (IMUs) have proven to be an ideal solution for tracking motion [69]. Healthcare practitioners can leverage this technology through progress-based applications such as motor function analysis and rehabilitation. However, despite emerging support for the effectiveness of IMU enabled mobile health applications, there is little emphasis on developing new visualizations for IMU motion trajectory [70, 33]. Erbaugh noted that visual feedback plays a pivotal role in one's ability to make correct movements. In the case of range of motion (ROM) assessment, visualization empowers both the doctor and patient to be more informed on one's motion that would otherwise be lost in a single quantitative number [16]. Therefore, by visualizing the motion trajectory, quantitative measures of motion can be evaluated while also assessing the quality of that movement.

Specifically, identifying the points in time when a patient's motion is deviating from an accepted reference trajectory can aid in understanding fatigue, ROM limitations, and various

movement qualities relevant to the exercise. Then, by leveraging analytical modeling, we can derive quantitative measures on the quality of movement to capture activity performance. As the motions that matter most will differ across diagnoses and applications, there will also be variance across the patients. The patient-to-patient variation here does not refer to the individualized preference of motion for increased quality of life, but rather the difference in severity of a patient’s underlying diagnosis. For example, the range of motor function impairment for children with SMA include those who can move independently and those whose only movement is a small flexion of the wrist. Comparative trajectory analysis is one method that can assess motion quality characteristics of wrist position and orientation, acceleration profile, and various angles.

In this work, we propose a new visualization framework that combines analytical modeling with comparative trajectory analysis to model IMU motion data into meaningful information related to the quantity and quality of motion.

### 3.2 Methodology



**Fig 3.1. MISTA Framework.** Our novel visualization framework can models deviations in motion trajectory.

In this section, we describe the steps of our framework, as seen in Figure 3.1. Note that our framework assumes the following conditions:

1. Trajectories are heterogeneous (reported in the same time reference)

- 
2. Movement is along one axis
  3. The signal is continuous

### **3.2.1 Feature Representation of Motion Data**

Prior to the use of our framework, we recommend an exploratory analysis of the features for the data. This may include box plots, kernel density estimation (KDE) plots, summary statistics, correlation heatmaps, etc. In doing so, this provides the foundation necessary to understand specific features of interest for comparing between trajectories.

### **3.2.2 Motion Alignment with Curve Matching**

The first step of our framework is to select an appropriate algorithm that compares two time series sequences through curve matching, as discussed in Section 2.3.2. Some common curve matching algorithms include ED, DTW, STS, TAM, and Longest Common Subsequence (LCSS). Though the occurrence of equal length trajectories is rare to happen in the wild, the length limitation for using ED as a distance metric can be overcome through the use of more robust distance measures, such as DTW. The algorithm selected must be able to compare the two sequences through ordered pairs and provide those ordered pairs as output.

### **3.2.3 Time Window and Statistical Feature Selection**

#### **Index**

Next, we use the indices of the ordered pairs as breakpoints for creating time windows. The user selects As the framework iterates through the ordered pairs, These linearly spaced points become the initial x-coordinate values in our graph. However, we found that it is more informative to show the elapsed time for a given movement of the query sequence. Using the timestamps of the query trajectory, we find the time delta between the time window indices, to show an elapsed time.

#### **Statistical Feature Selection**

Since the time windows are an aggregation of multiple points, the user must select a statistical feature to represent each time window, which will become the Y-coordinate values. Based

---

on the exploratory analysis, this may drive the decision of which feature to select, though, simple features such as maximum, minimum, mean, and standard deviation are good places to start. Other values can also be selected, such as the first, last, or median value of the time window as well.

### 3.2.4 Trajectory Divergence Analysis

The final step of our framework involves comparing the time window features from the query sequence to the reference sequence by either computing the absolute difference or through the use of a distance metric. This will determine the dot size. Figure 3.2 (a) depicts the absolute difference between the mean values of sample set of trajectories.

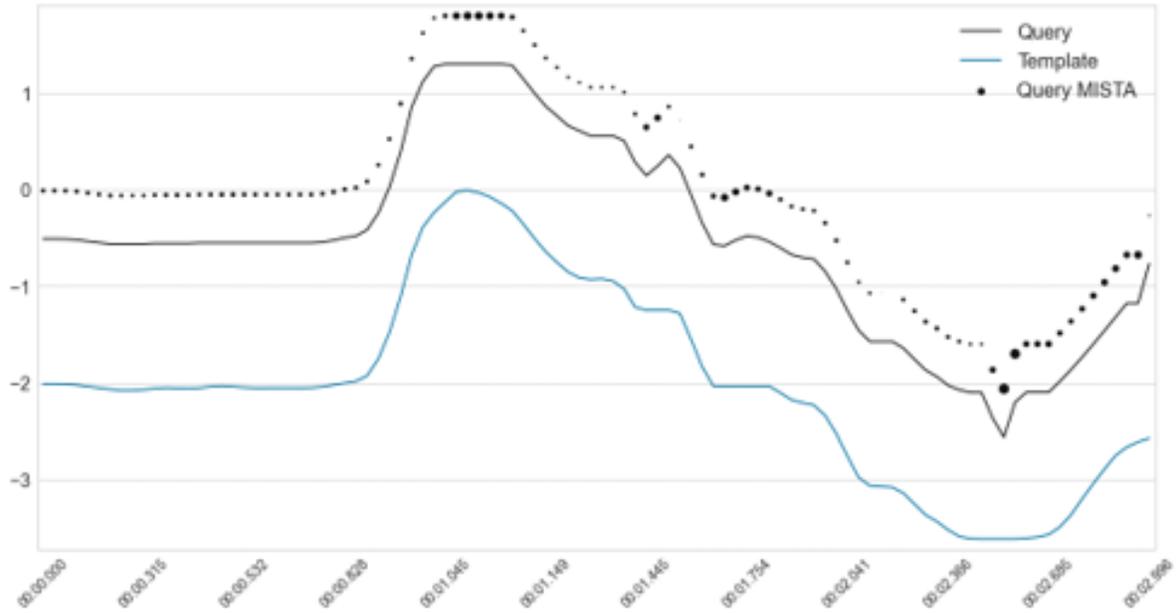
#### Statistical Tests

Furthermore, a statistical test can also be used to adjust the dot size if the difference is found to be significant at the desired alpha level. Figure 3.2 (b) illustrates how the use of a statistical test, such as the one-way ANOVA, can influence the dot size, providing a clearer picture of where the most important deviations occur along the trajectory.

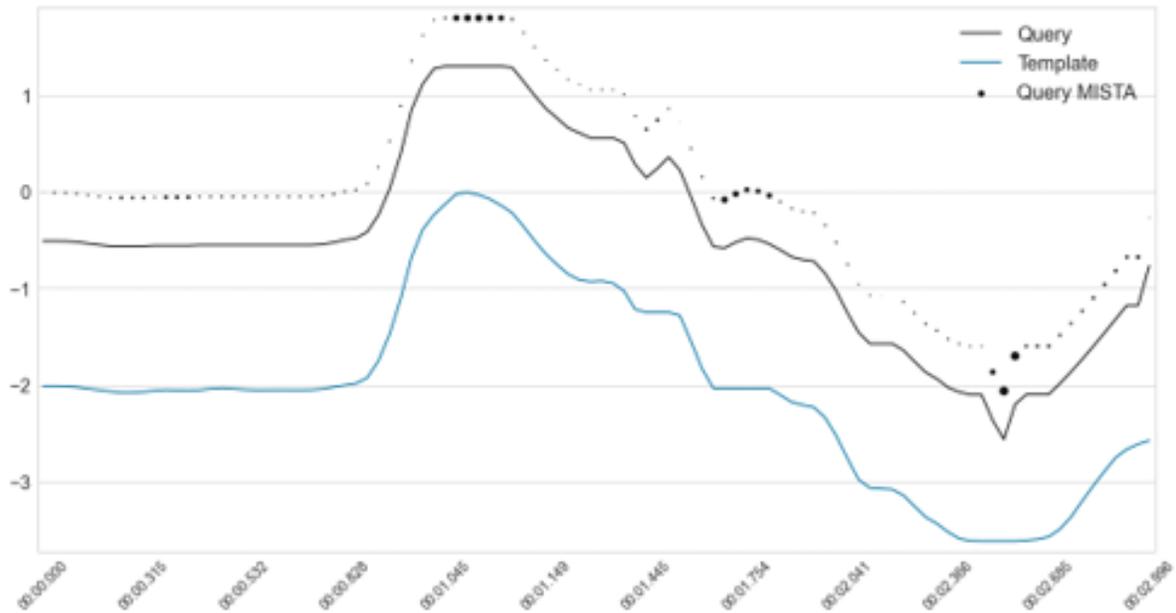
### 3.2.5 Visualization

After completing the steps above, we can now plot the aligned query and template sequences with the Matching Index Statistical feature Trajectory Analysis (MISTA) dots. To achieve the best results, this process may need to be repeated for several features as the differences between the sequences will vary. MISTA will aid in the discovery of distinct points in time when the two trajectories are not in sync.

To demonstrate how our framework can be used, we use gait analysis case study. Imagine there is a participant walking on a treadmill and we place a sensor on each thigh. From the accelerometer data, we extract one full gait cycle where we identify when the toe come off the ground, swings forward, makes contact with the ground, and becomes flat just before the toe comes off the ground again. For this case study, we purposely kept the non-dominant foot on the ground longer at two time points, to introduce distinct differences that should be



(a) Relative Divergence

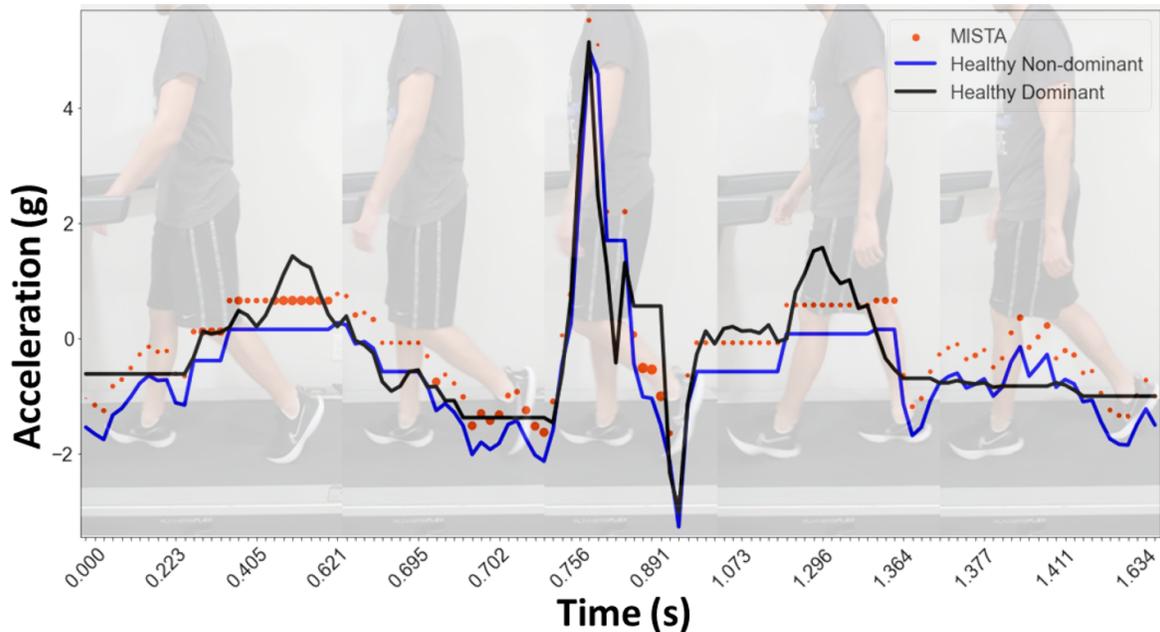


(b) Statistically Significant Divergence

**Fig 3.2. Divergence Analysis.** The dot size can be changed depending on the user's choice for comparing the statistical feature. The top plot (a) shows the dot size relative to the absolute difference between the time windows of the template and query trajectories.

The bottom plot (b) illustrates larger dots for the time windows that are deemed statistically significant after applying a statistical test to compare the time windows. Here, we show the use of a one-way ANOVA test with  $\alpha = 0.05$  to increase the dot size by 3x.

captured. The goal of the case study was to compare the symmetry between the acceleration profile of the participant’s dominant and non-dominant leg.



**Fig 3.3. MISTA Gait Example.** As an example, we apply MISTA to a healthy participant’s gait cycle as they walk on the treadmill.

The scenario mentioned is seen in Figure 3.3, which shows phases from healthy participant 1’s gait cycle with the corresponding MISTA framework plot that compares the symmetry between the acceleration of their dominant and non-dominant leg. The black line is the template or reference trajectory (participant’s dominant leg) while the blue line is the query trajectory (participant’s non-dominant leg). The MISTA plot in Figure 3.3 uses DTW to align the trajectories since the lengths were unequal. After extracting the matched indices to form the time windows, we selected the maximum value as the statistical feature, which becomes the new y-coordinate values. Then, we found the absolute difference of the max bin values between the query trajectory and the template trajectory. As an additional measure, we also computed the one-way ANOVA test and adjusted the dot size three times larger if the difference was significant at an alpha of 0.05. Finally, we plot the new y values along the same range as the full trajectories, and use the new bin distance measure output as the dot size of each bin. The dot size indicates where the largest deviations are along the query

---

trajectory.

By applying our framework to the acceleration profile of healthy participant 1's gait, we are able to visualize and extract the exact timing of deviations from symmetry. From Figure 3.3, we can see that the participant's acceleration profile is generally the same for each leg, with the exception of two peaks from the dominant leg (template). The first peak around 0.493 seconds is in relation to the toe lift phase while the second peak around 1.2 seconds is the leg coming forward after the foot becomes flat with the ground. The acceleration and deceleration during the swing phase are nearly identical. The MISTA plot was able to correctly identify the two time points that were purposely performed different.

### **3.3 Discussion**

MISTA provides researchers and physicians the ability to assess quantitative outcomes paired with measurable metrics on the quality of the movements, providing informative data for rehabilitation interventions and monitoring recovery. Our framework could be applied to a variety of applications, like shape detection, signal processing, other pattern recognition problems. Healthcare practitioners can leverage MISTA through progress-based applications such as motor function analysis and rehabilitation. Specifically, identifies the points in time when a patient's motion is deviating from an accepted reference trajectory, which can aid in understanding fatigue, ROM limitations, and various movement qualities relevant to the exercise.

## Chapter 4

# Validation through Human Motion Capture with Wearable Sensors

### 4.1 Leveraging Wearable Sensors to Detect and Assess Rehabilitation Progress

#### 4.1.1 Background and Motivation

Anterior cruciate ligament (ACL) rupture is a common and costly musculoskeletal injury, affecting 250,000 people in the United States and costing over 3 billion annually [71, 72, 73]. Successful return-to-sport (RTS) following ACL reconstruction (ACLR) surgery relies on the ability to assess muscle function during rehabilitation. Research has shown that young athletes, especially females, are more likely to experience an ACL injury and are more prone to sustaining a contralateral ACL injury after the first occurrence [74]. Additionally, studies have also found that many athletes' timeline for return-to-sport (RTS) is too soon [75]. Therefore, the ability to detect and track subtle movement asymmetries in native environments is essential to track progress early and throughout the course of rehabilitation and recovery.

To aid in the decision making process for RTS, clinicians rely on performance evaluations and their associated outcomes. Many physical movement evaluations are performed

---

using video-graphic or optoelectronic systems integrated with hardware and software systems. These kinds of systems (e.g., Vicon) involve applying markers to the subject, setting up and calibrating the reference frame, video or marker recording, digitization, transformation, smoothing and normalization [76]. These evaluations require elaborate preparation and set-up, rendering this type of evaluation to be cost and time prohibitive for multiple evaluations. However, the gold standard evaluation uses an isokinetic dynamometer to assess the quadriceps femoris (QF) strength symmetry. Isokinetic strength testing provides a limb strength symmetry (LSI) ratio of the involved leg to the uninvolved leg, where 90% or better is considered a benchmark for progression of rehabilitation toward RTS. Further, these evaluations are limited to controlled and unrealistic laboratory environments that have little similarity to “in-field” activity or sport exposure because analysis is limited to one type of movement patterns during relatively few steps.

Recent advances in ubiquitous sensing have the potential to fill these gaps. Monitoring physical activity with wearable sensors provides more portability than in-clinic/lab instruments and can be worn over an extended period of time in an unrestricted environment. Sensors such as accelerometers, gyroscopes and surface electromyography (sEMG) can be leveraged to assess gait, mobility, and muscle activity characteristics in athletes’ natural environments. Accelerometers and gyroscopes have been used extensively to monitor gait and assess athlete performance [7, 3]. The sEMG signal is a measurement of electrical current generated in muscles during contraction. Previously, the cost, complexity, and usability of sEMG devices have prevented clinicians from realizing the benefits of muscle activity information for assessing their patients. However, recent technological advances and the availability of more sEMG solution have lowered the barriers to entry. Surface Electromyography (sEMG) devices have been shown to recognize hand and arm gestures while worn on the forearm by recording electrical impulses stemming from muscular contractions through the use of surface electrodes [22]. Often times, sEMGs are paired with accelerometers or IMUs to better capture motion. The limitations of these devices are similar to that of

---

smartwatches, where they are generally worn on one limb at a time and do not have very high sampling rates.

Recognizing the applicability of ubiquitous sensing, the UVA Exercise and Sport Injury Lab sought to leverage advanced sensing technologies and innovative analytics to answer the following research questions:

**Research Question 1:** Can body worn sensors be used to differentiate between healthy and patient populations during rehabilitation?

**Research Question 2:** How can body worn sensors inform rehabilitation programs on the status of patient progression?

The study will seek to answer the research questions above through signal feature analysis. We will first describe the study design for the ACLR dataset, and extract a set features to explore with descriptive analysis. We will then design classification models using the extracted features to differentiate between the healthy and patient cohorts. We will then present a correlation analysis using motion based LSI features against the LSI values provided by the isokinetic dynamometer. Ultimately, this study serves as the initial step in determining the feasibility of using body worn sensors to aid in treatment decision making regarding the safety of patients RTS after severe knee joint injuries.

#### 4.1.2 Study Design

##### Subjects

Our study involved twenty-two participants, with twelve patients (first-time ACL reconstruction and free from complication) and ten healthy subjects sex: 4F/6M, age: 22–35 yrs, height: 1.6–1.8 m, weight: 61.2–99.3 kg). Our analysis will focus on twenty-one participants as one patient’s data was corrupted and unrecoverable. Patients in the study are referred by their surgeon at the UVA Hospital while healthy participants are young athletes and students attending UVA. Each patient completes two self-reported knee function assessments, the International Knee Documentation Committee subjective knee form (IKDC) and the

---

Knee injury and Osteoarthritis Outcome Score (KOOS). Both assessments have been found to be reliable measures for patients who have undergone ACLR surgery [65, 66, 67, 68]. The IKDC is based on 0-to-100 scale, with higher scores translating to more knee function while the KOOS is the combination of five subscales: pain, symptoms, activities of daily living, sport and recreation activities, and quality of life. Each subscale has questions scored between 0 and 5, which are then aggregated on a 0-to-100 scale, with higher scores indicating less knee issues.

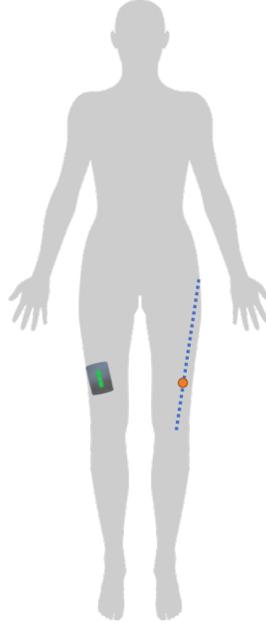
### **Experimental Protocol**

The study involves a walking gait analysis, a muscle strength assessment using an isokinetic dynamometer (Biodex Medical Systems Inc., NY, USA), followed by hopping tests, if able. The isokinetic dynamometer controls the speed and resistance for a given range of motion to evaluate muscle strength. Our study involves the placement of a Delsys Trigno™ sensor (Delsys Inc., MA, USA) on each quadriceps femoris to monitor these physical activities 3-6 months post-surgery. The sensor location was determined by first measuring the distance from the greater trochanter to the superior-lateral pole of the patella. Then, the sensor was placed on the vastus lateralis in line with the muscle fiber orientation at the distal 1/3 of the measured distance, as seen in Fig 4.1.

Each subject began by walking on the treadmill for five minutes, and then performed a set of isokinetic testing exercises at 90 and 180 deg/s. From the exercises, the muscle strength is translated into a Limb Symmetry Index (LSI), calculated by Eq 4.1. The LSI ratio has been found to be a reliable measure for assessing asymmetries [77].

$$LSI = (involved/uninvolved) \tag{4.1}$$

Depending on a patient's progress and pain tolerance, more exercises could be performed, such as the single leg hop. However, we did not include this exercise in our analysis since most of the patients were unable to perform the movement.

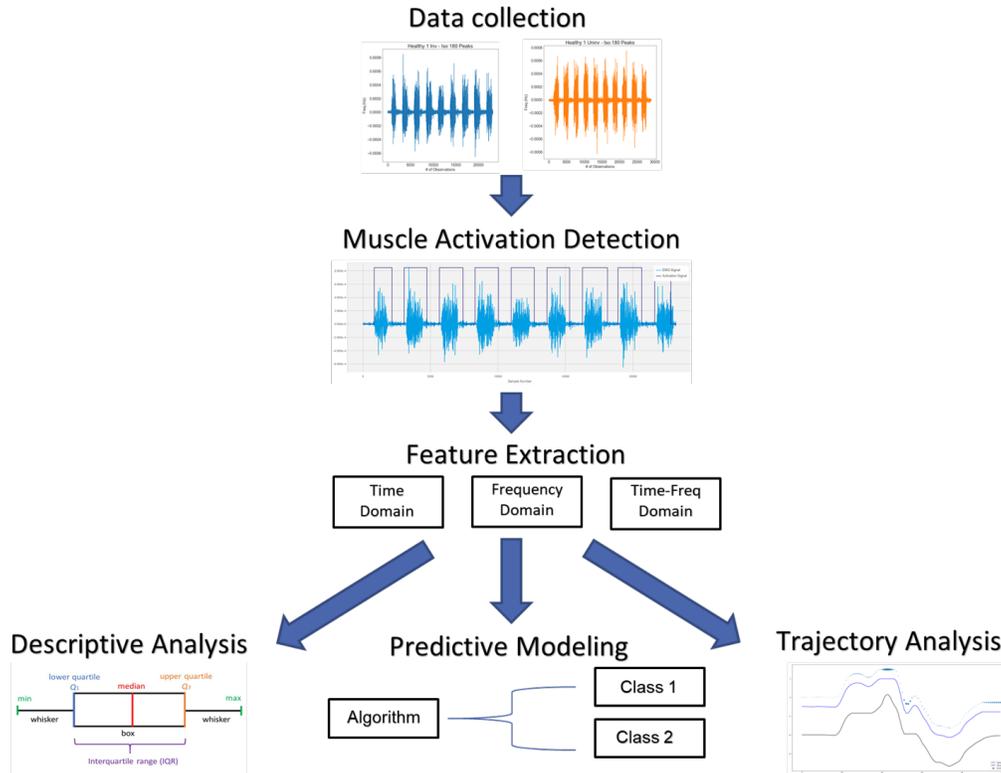


**Fig 4.1. Sensor placement.** Delsys Trigno sensors were applied bilaterally at a measured distance on the vastus lateralis for tracking gait kinematics and quadriceps muscle activations.

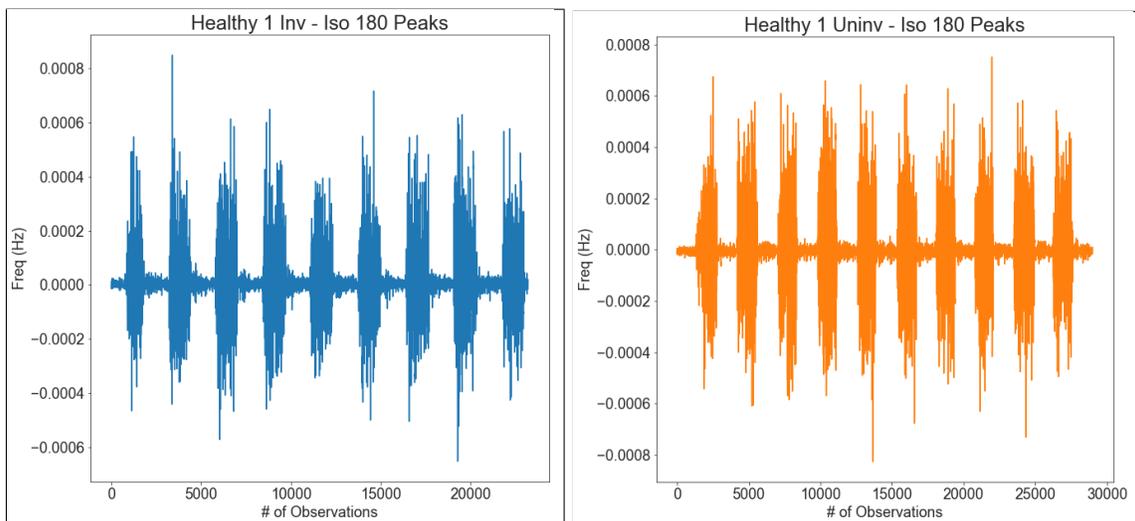
### 4.1.3 Methodology

To answer the first question, we use the methodology shown in Fig 4.2. It follows standard HAR approaches, beginning with data collection and data processing. Next, there is muscle activation detection followed by feature extraction. Then, the features are summarized with descriptive analysis modeled with machine learning algorithms for classification.

The accelerometer and sEMG data were sampled at 148 Hz and 1926 Hz, respectively. Both sensors underwent processing and filtering through Delsys' proprietary software. To limit sensor noise, the placement location was first shaved and cleaned on each participant prior to placement. Then each sensor was wrapped with medical tape to limit movement over the duration of the exercises. Further data processing and all analyses were accomplished with python using custom scripts. The sEMG data was reprocessed through a band-pass filter between 10-500 Hz to remove any remaining noise. Fig 4.3 depicts the sEMG muscle activations for healthy subject 1 during the iso 180 exercise, where blue represents their involved leg and orange is their uninvolved leg.

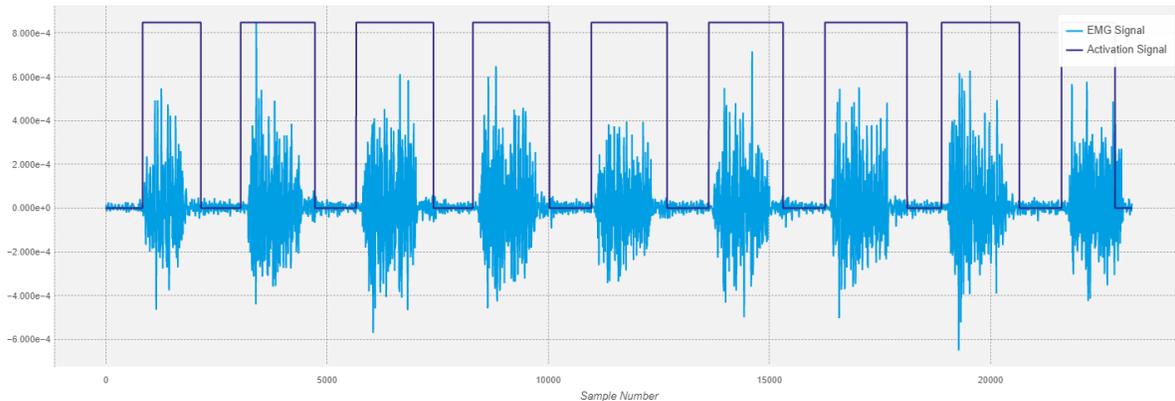


**Fig 4.2. Methodology breakdown.** Our methodology begins with data collection, followed by muscle activation detection, and feature extraction. Then, we provide a descriptive analysis of the motions, apply predictive modeling to classify the motions into their respective cohorts, and then apply our MISTA framework for trajectory analysis of limb symmetry.



**Fig 4.3. Iso 180 EMG data** The muscle activations for healthy subject 1 during the iso 180 motion are shown with the involved leg on the left and the uninvolved leg on the right.

Focusing on the isokinetic exercises, we performed our analysis using the sEMG signal data since the placement of the sensor would cause there to be no acceleration during the specific exercises. The cyclical pattern found in the sEMG data suggested that the individual muscle activations could be extracted for comparative analysis. Looking at each subject’s leg, we separated out the muscle activations during each exercise, as shown in Fig 4.4.



**Fig 4.4. Muscle activation separation.** The muscle activations of healthy subject 1’s involved leg during the iso 180 exercise are separated into individual partitions for comparative analysis.

Then, we extracted the features described in Table 4.1 below for each activation. These features were chosen based on previous work [78, 79]. Next, we found the average of each feature to represent the subject’s performance for a given leg on each exercise. Since each leg was measured independently, we also computed LSI ratios for each feature.

Within each exercise, we perform the 2-sided Mann-Whitney-U statistical test with a significance level equal to 0.05, to test whether the mean values of each class were equal across each extracted feature for six scenarios: healthy vs patient, involved vs uninvolved, healthy involved vs healthy uninvolved, patient involved vs patient uninvolved, healthy involved vs patient involved, and healthy uninvolved vs patient uninvolved.

---

**Table 4.1. Extracted features**

<b>Feature</b>
Variance
Root Mean Square (RMS)
Integrated EMG (IEMG)
Mean Absolute Value (MAV)
Log Detector (LOG)
Wavelength (WL)
Average Amplitude Change (ACC)
Difference Absolute Std Dev Value (DASDV)
Willison Amplitude (WAMP)
Myopulse Percentage Rate (MYOP)
Frequency Ratio (FR)
Mean Power (MP)
Total Power (TP)
Mean Frequency (MNF)
Median Frequency (MDF)
Peak Frequency (PKF)
H-Wavelet (WENT)

This table lists all the features extracted from each muscle activation that will be used for comparative analysis.

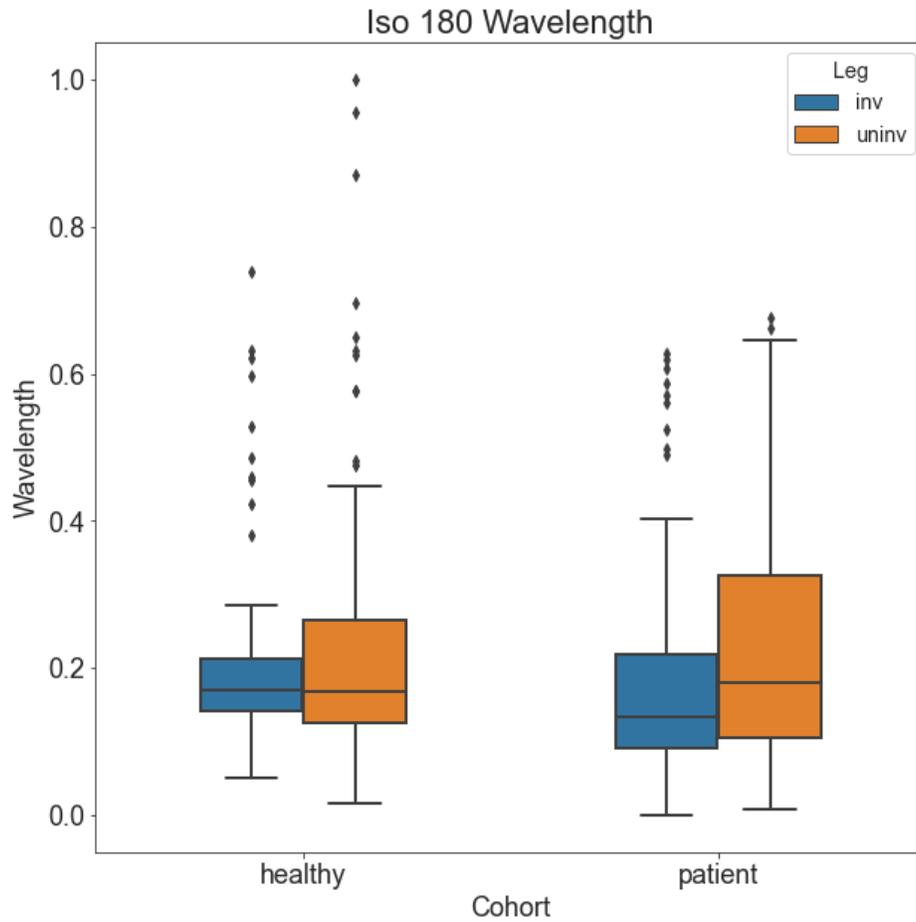
#### **4.1.4 Results**

##### **4.1.4.1 Descriptive Analysis**

###### **Iso 180**

Beginning with the iso 180 exercise, we notice differences in the extracted features between the cohorts and the subjects' legs. These differences are apparent when looking at the boxplots of the individual features. For example, in Fig 4.5, we see that the median for the healthy involved and uninvolved legs are very similar, despite their interquartile range being uneven. For the patient cohort, we see the uninvolved leg has a higher median value and a much higher upper bound for the 3rd quartile.

To analyze all the remaining features, we applied a scaling function and plotted the boxplots, as seen in Fig 4.6. Only five of the eighteen features were found to have correlation values less than 0.75, which included VAR, WL, ZC, WAMP, MYOP and FR. From Fig



**Fig 4.5. Wavelength (WL) feature boxplot for iso 180.** For the iso 180 exercise, the median WL values of the involved and uninvolved legs within the healthy cohort are nearly identical while the median WL value for the patients' uninvolved leg is shown to be much higher than their involved leg.

---

4.6, we observe that the median value of the uninvolved leg for both healthy and patients is generally higher in the time domain than that of the involved leg. In the frequency domain, the MNF, MDF, PKF, and WENT features are all inversely related between the cohorts. With the exception of the WENT feature, the patient’s involved leg has the lowest median value across all the features.

To determine whether the differences between the features are statistically significant for the 6 scenarios previously mentioned, we look to Table 4.2 for the results of the Mann-Whitney-U test. In this table, the values highlighted in gray represent the features found to be statistically significant in each scenario. From this table, we can clearly see that as a whole, the patients’ involved leg performs differently than the uninvolved leg for the iso 180 exercise.

### **Iso 90**

Next, we look at the iso 90 exercise. In Fig 4.7 we see that the median and interquartile range for the healthy involved and uninvolved legs are very close. For the patient cohort, we see the uninvolved leg has a higher median value, yet the interquartile range is smaller than the involved leg.

Assessing the remaining features, Fig 4.8 displays the scaled feature boxplots for this exercise. The set of uncorrelated features were almost identical to the iso 180 exercise, with VAR, WL, ZC, WAMP, MYOP, FR and PKF having correlation values less than 0.75. With the exception of the ZC and MYOP features, the healthy uninvolved leg has a higher median value than the involved leg for all time domain features. The patient uninvolved leg had a higher median than the involved leg, in all the features except FR. This observation is unlike that of the iso 180 exercise, where features had an inverse relationship between the cohorts. Instead, the two cohorts performed similarly, with the exception of the PKF feature.

The results of the statistical tests for the iso 90 features are shown in Table 4.3. The values highlighted in gray represent the features found to be statistically significant in each scenario. Similar to the iso 180 exercise, there is a clear indication that the patient involved

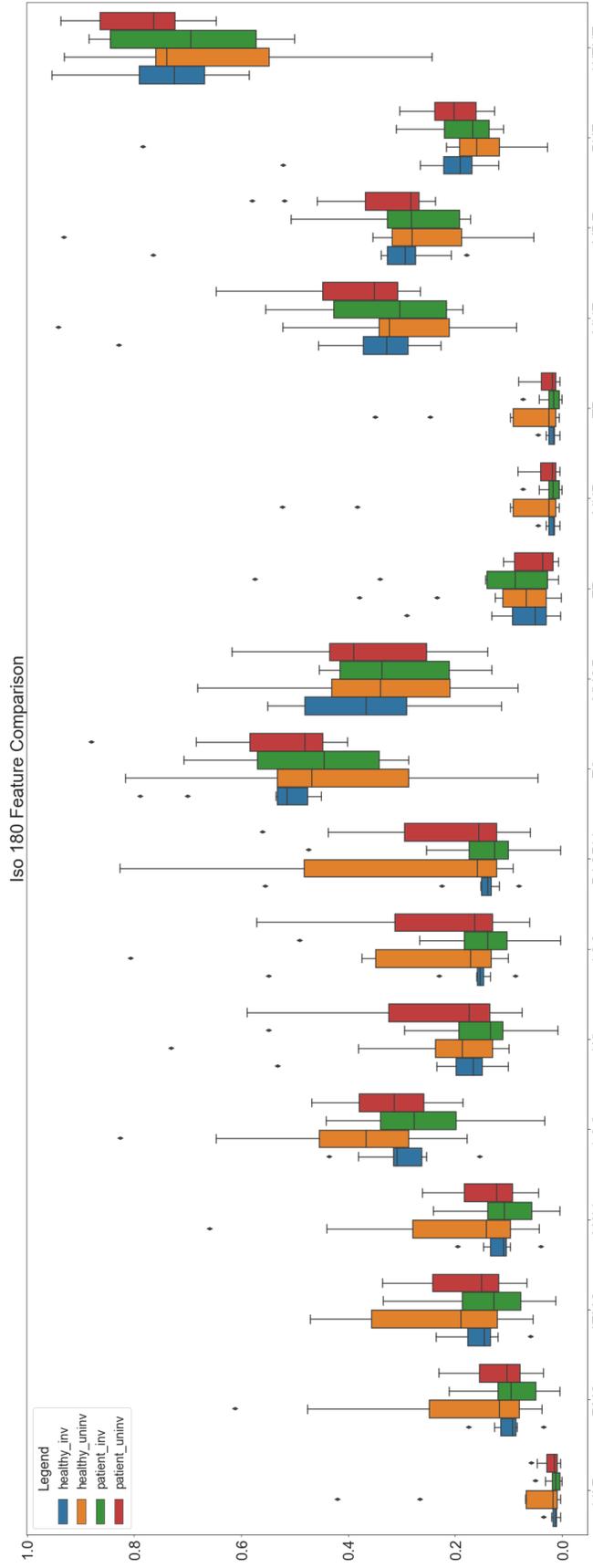


Fig 4.6. Iso 180 feature comparison.

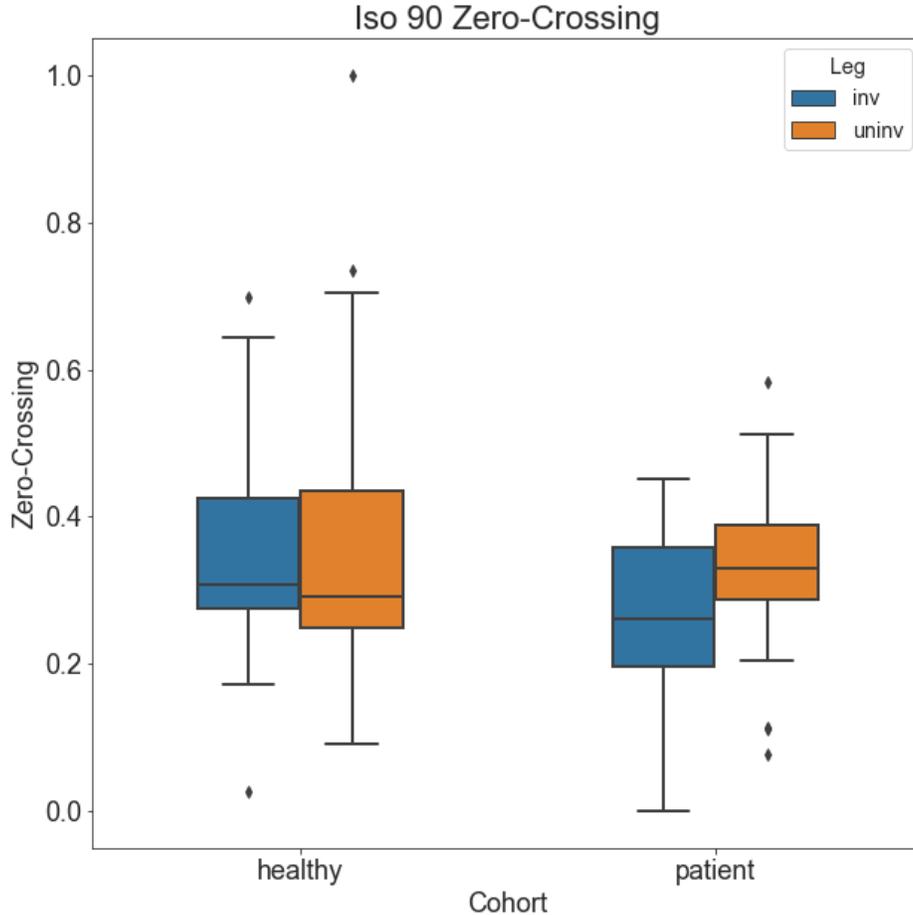
**Table 4.2.** Iso 180 Mann-Whitney-U Statistical Tests

	<b>H v P</b>	<b>Inv v Uninv</b>	<b>Hinv v Huninv</b>	<b>Pinv v Puninv</b>	<b>Hinv v Pinv</b>	<b>Huninv v Puninv</b>
<b>VAR</b>	0.3470	0.0171	0.2684	0.0348	0.4111	0.5881
<b>RMS</b>	0.3470	0.0171	0.2684	0.0348	0.4111	0.5881
<b>IEMG</b>	0.1698	0.0108	0.1693	0.0225	0.1369	0.5386
<b>MAV</b>	0.3736	0.0066	0.1768	0.0243	0.5009	0.5328
<b>LOG</b>	0.3139	0.0009	0.0194	0.0141	0.4705	0.3778
<b>WL</b>	0.0513	0.0110	0.9618	0.0016	0.0019	0.9473
<b>ACC</b>	0.1203	0.0094	0.9872	0.0024	0.0151	0.8984
<b>DASDV</b>	0.1883	0.0154	0.9841	0.0038	0.0250	0.7568
<b>ZC</b>	0.0249	0.3940	0.0352	0.0034	0.0001	0.4866
<b>WAMP</b>	0.0014	0.7002	0.7591	0.7396	0.0247	0.0234
<b>MYOP</b>	0.7311	0.7628	0.0232	0.1083	0.0637	0.0228
<b>FR</b>	0.2345	0.0764	0.0412	0.0000	0.0003	0.0593
<b>MNP</b>	0.2324	0.0119	0.2548	0.0219	0.2024	0.5641
<b>TP</b>	0.1911	0.0192	0.2878	0.0285	0.1749	0.5251
<b>MNF</b>	0.5145	0.0568	0.2701	0.0005	0.0170	0.1333
<b>MDF</b>	0.3084	0.0807	0.7130	0.0041	0.0440	0.5639
<b>PKF</b>	0.4014	0.5915	0.1825	0.0448	0.0273	0.3035
<b>WENT</b>	0.7299	0.0057	0.6007	0.0001	0.0078	0.0690

The results of the Mann-Whitney-U statistical tests for the Iso 180 exercise are presented, comparing the mean feature values (bold left column) between the 6 scenarios or subsets shown in the bold on the top row. Highlighted values indicate where the tests are found to be statistically significant at an alpha of 0.05, thus rejecting the null hypothesis that the mean values for the two classes are equal.

leg is performing differently than the involved leg.

As we examine the features across the cohorts in Figs 4.6 and 4.8, we can see the same trends arise between the cohorts for each exercise. Generally, the median values within the healthy cohort are almost the same, despite the interquartile range of the healthy uninvolved leg being larger than the involved leg. This observation is validated by the statistical test results shown in Tables 4.2 and 4.3, where the statistical tests comparing the healthy subjects' legs for WL (iso 180) and ZC (iso 90) are not significant and the tests for the patients' are highlighted as significant. Another trend found by the boxplot analysis is that the upper bound of the frequency domain features were usually higher for involved leg of the healthy



**Fig 4.7. Zero-Crossing (ZC) feature boxplot for iso 90.** Here we see that within the healthy cohort, the median ZC value is very similar between the involved and uninvolved leg. However, there appears to be a clear difference between the median ZC values for the patient cohort.

subjects, which is opposite of the time domain features and contradictory for the patients. In validating our hypotheses that the healthy subjects would show no difference between their legs while patients would, we aimed to capitalize on those differences through classification.

#### 4.1.4.2 Predictive Modeling of Limb Symmetry

We propose to leverage supervised machine learning algorithms to determine how well the derived sEMG features can distinguish between the cohorts as well as between the involved and uninvolved legs. Following previous work from [80] et al, some of the top ML classifiers include random forest (RF), linear discriminant analysis (LDA), gradient boosting machine (GBM), and k-nearest neighbor (KNN). Due to the limited size of the dataset, we used

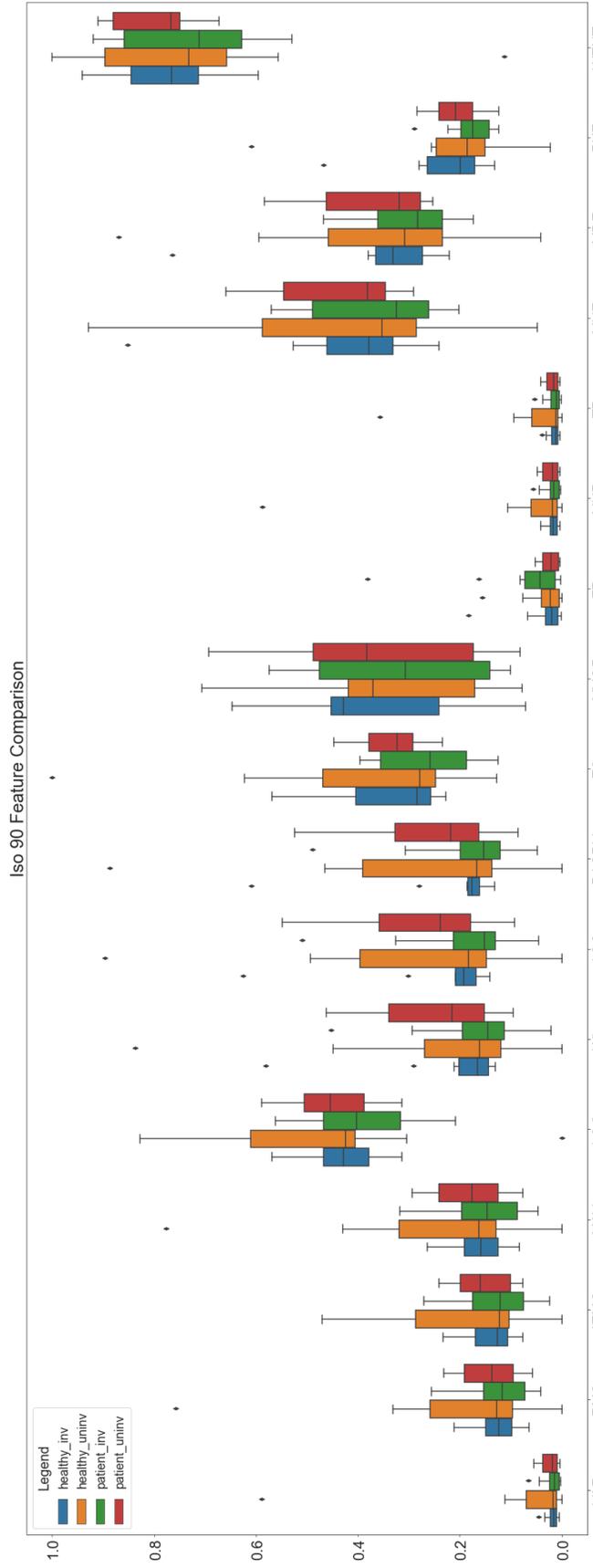


Fig 4.8. Iso 90 feature comparison.

**Table 4.3.** Iso 90 Mann-Whitney-U Statistical Tests

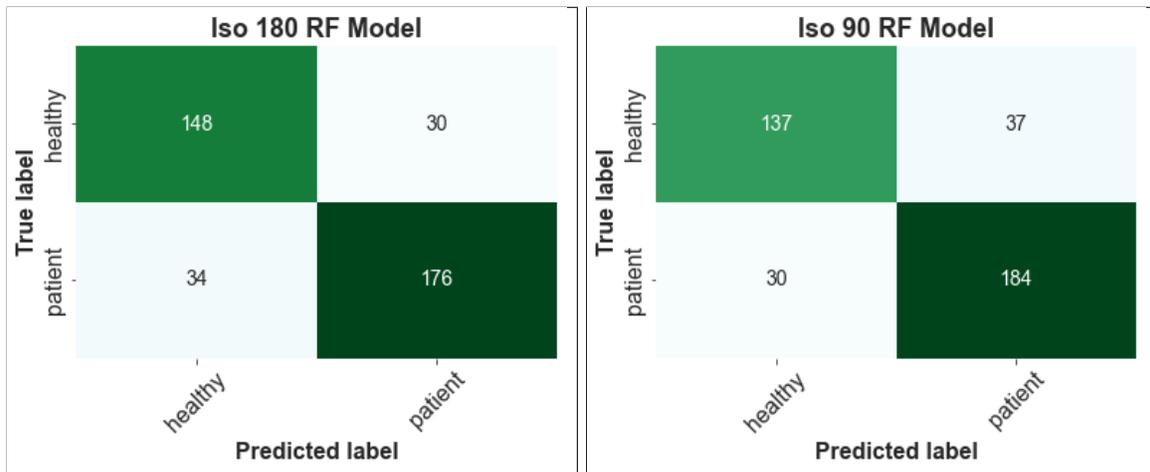
	<b>H v P</b>	<b>Inv v Uninv</b>	<b>Hinv v Huninv</b>	<b>Pinv v Puninv</b>	<b>Hinv v Pinv</b>	<b>Huninv v Puninv</b>
<b>VAR</b>	0.9250	0.0002	0.0444	0.0023	0.5590	0.9849
<b>RMS</b>	0.9250	0.0002	0.0444	0.0023	0.5590	0.9849
<b>IEMG</b>	0.9066	0.0012	0.1781	0.0045	0.4201	0.8506
<b>MAV</b>	0.8250	0.0001	0.0317	0.0012	0.3965	0.9337
<b>LOG</b>	0.5254	0.0000	0.0115	0.0001	0.1796	0.7647
<b>WL</b>	0.4880	0.0007	0.9510	0.0000	0.0100	0.2482
<b>ACC</b>	0.2167	0.0000	0.9926	0.0000	0.0005	0.2434
<b>DASDV</b>	0.4836	0.0003	0.7557	0.0000	0.0061	0.1913
<b>ZC</b>	0.0808	0.0292	0.1342	0.0000	0.0000	0.0627
<b>WAMP</b>	0.0863	0.4165	0.7536	0.1381	0.6516	0.0407
<b>MYOP</b>	0.2524	0.7739	0.1308	0.3290	0.0371	0.5750
<b>FR</b>	0.0171	0.0889	0.0221	0.0000	0.0000	0.1597
<b>MNP</b>	0.5201	0.0008	0.0587	0.0069	0.4153	0.6435
<b>TP</b>	0.9739	0.0006	0.1047	0.0040	0.5205	0.8529
<b>MNF</b>	0.3223	0.0861	0.0121	0.0000	0.0009	0.0370
<b>MDF</b>	0.2161	0.0672	0.1150	0.0001	0.0035	0.2340
<b>PKF</b>	0.0190	0.2024	0.2807	0.0035	0.0006	0.9105
<b>WENT</b>	0.3933	0.0627	0.0175	0.0000	0.0003	0.0178

This table displays the results of the Mann-Whitney-U statistical tests for the Iso 90 exercise, comparing the mean feature values (bold left column) between the 6 scenarios or subsets shown in the bold on the top row. Highlighted values indicate where the tests are found to be statistically significant at an alpha of 0.05, thus rejecting the null hypothesis that the mean values for the two classes are equal.

leave-one-out-cross-validation to aid in assessing the performance of each the classification models.

A RF classification model uses multiple sub-samples of the dataset to fit a set of decision trees that are then averaged for the output[81]. GBM is similar to random forest, where the algorithm is built on decision trees, however, the algorithm learns from the previous models built, to reduce the overall error [82]. LDA is an algorithm that focuses on dimensionality reduction by transforming the data into a set of component axes that maximize the variance of the data while also maximizing the separation between multiple classes [83]. KNN is another supervised learning algorithm designed for classification that aims to predict the

associated class by selecting the 'K' nearest points according to a specified distance metric [84].



**Fig 4.9. Classification confusion matrices.** Confusion matrices for the top classification models in the healthy versus patient scenario. The left is the random forest model for the iso 180 exercise while the right is the random forest model for the iso 90 exercise.

We also present the top performing classification of each exercise using a confusion matrix. Confusion matrices offer insight into how each algorithm classified the samples and which classes were predicted most correct. Two scores that can be generated from the confusion matrix are precision and recall. Precision refers to the ratio of true positives over the total of true positives and false positives. Recall refers to the ratio of true positives over the total of true positives and false negatives. Combining the precision and recall into a single metric yields the F1-score.

For the iso 180 exercise, the classification accuracy for each algorithm is shown in Table 4.4 below. Given that the random forest model achieved the highest classification accuracy, the confusion matrix is shown on the left in Fig 4.9. From this chart, we see that the model performed similar for each class, with a F1-score of 0.82 for healthy and 0.85 for patient.

For the iso 90 exercise, the classification accuracy for each algorithm is shown in Table 4.5 below. For the iso 90 exercise, the random forest model also achieved the highest classification accuracy. From Fig 4.9 (right), we see that the model performed similar for each class, with

**Table 4.4. Classification Accuracy for Iso 180**

	<b>RF</b>	<b>GBM</b>	<b>KNN</b>	<b>LDA</b>
<b>H v P</b>	83.5% $\pm$ 0.4	81.4% $\pm$ 0.4	71.4% $\pm$ 0.5	67.0% $\pm$ 0.4
<b>Inv v Uninv</b>	80.8% $\pm$ 0.4	78.8% $\pm$ 0.4	66.7% $\pm$ 0.5	59.9% $\pm$ 0.5
<b>Hinv v Huninv</b>	83.1% $\pm$ 0.4	81.5% $\pm$ 0.4	69.1% $\pm$ 0.5	62.4% $\pm$ 0.5
<b>Pinv v Puninv</b>	83.8% $\pm$ 0.4	86.2% $\pm$ 0.4	79.5% $\pm$ 0.4	67.1% $\pm$ 0.5
<b>Hinv v Pinv</b>	88.1% $\pm$ 0.3	85.1% $\pm$ 0.4	77.8% $\pm$ 0.4	76.8% $\pm$ 0.4
<b>Huninv v Puninv</b>	83.0% $\pm$ 0.4	79.4% $\pm$ 0.4	77.8% $\pm$ 0.4	73.7% $\pm$ 0.4

a F1-score of 0.80 for healthy and 0.85 for patient.

**Table 4.5. Classification Accuracy for Iso 90**

	<b>RF</b>	<b>GBM</b>	<b>KNN</b>	<b>LDA</b>
<b>H v P</b>	82.7% $\pm$ 0.4	80.7% $\pm$ 0.4	70.9% $\pm$ 0.5	64.7% $\pm$ 0.5
<b>Inv v Uninv</b>	83.5% $\pm$ 0.4	80.8% $\pm$ 0.4	67.9% $\pm$ 0.5	65.2% $\pm$ 0.5
<b>Hinv v Huninv</b>	87.4% $\pm$ 0.3	81.0% $\pm$ 0.4	69.5% $\pm$ 0.5	60.9% $\pm$ 0.5
<b>Pinv v Puninv</b>	87.9% $\pm$ 0.3	86.9% $\pm$ 0.3	79.9% $\pm$ 0.4	78.5% $\pm$ 0.4
<b>Hinv v Pinv</b>	88.1% $\pm$ 0.3	86.1% $\pm$ 0.3	82.0% $\pm$ 0.4	78.9% $\pm$ 0.4
<b>Huninv v Puninv</b>	83.0% $\pm$ 0.4	82.0% $\pm$ 0.4	76.8% $\pm$ 0.4	70.1% $\pm$ 0.5

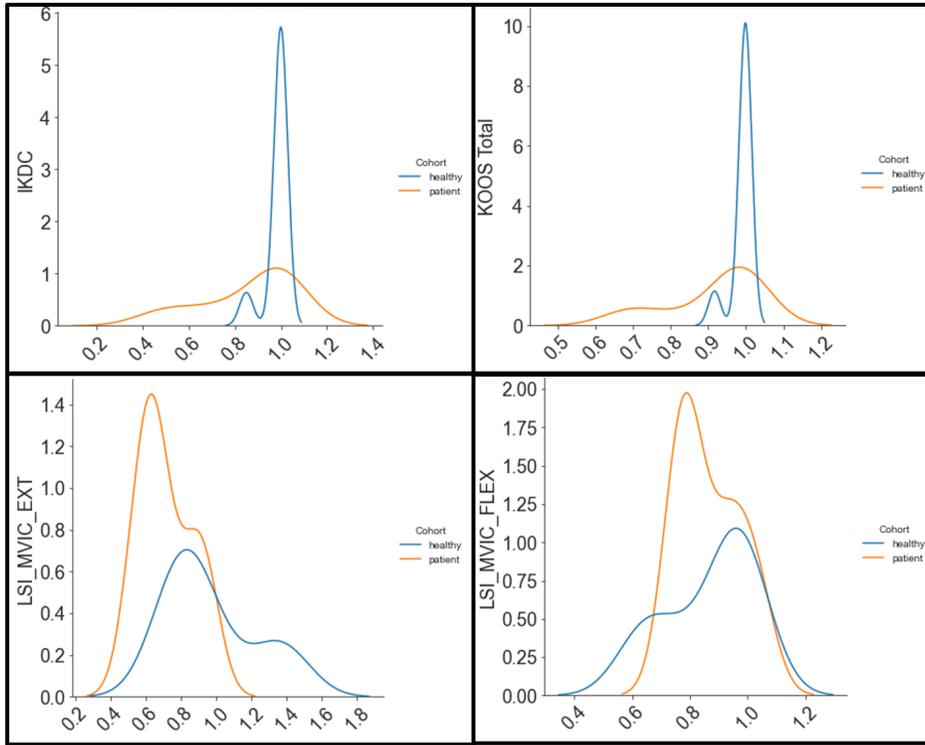
The two isokinetic movements had very similar accuracy predicting across the six scenarios. Therefore, we also did a combined classification task, using both exercises to predict the same scenarios in Table 4.6. For the healthy versus patient scenario, we see an increase in accuracy across all 4 models, especially in the KNN algorithm. The KNN algorithm showed improvement in all six scenarios, which suggests that more data was needed to improve the results. However, the increase in accuracy was not as comprehensive for the remaining scenarios. The final scenario, healthy uninvolved versus patient uninvolved, showed the second most improvement, with changes in the random forest, KNN, and LDA algorithms.

#### 4.1.4.3 Limb Symmetry Index (LSI) Comparison

To answer the second research question of this study, we computed our own set of LSI values from the extracted features of each exercise, and compared them to the LSI values provided by the isokinetic dynamometer. We interpreted the LSI values with a kernel density estimation (KDE) plot, which allows us to visualize the underlying distribution.

**Table 4.6. Classification Accuracy for Combined Exercises**

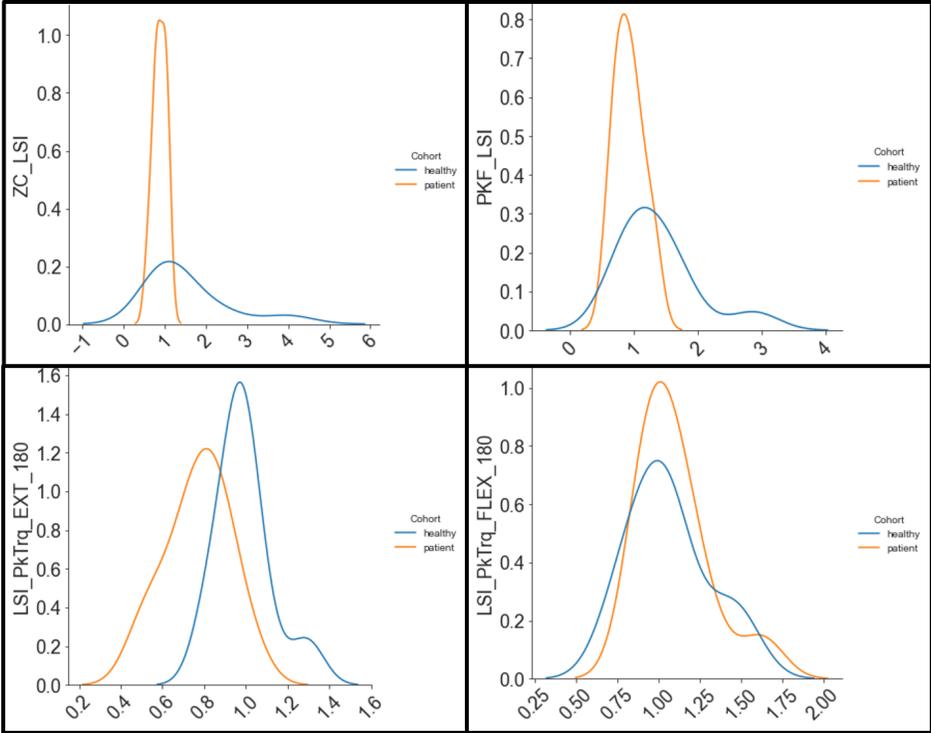
	<b>RF</b>	<b>GBM</b>	<b>KNN</b>	<b>LDA</b>
<b>H v P</b>	84.3% $\pm$ 0.4	82.1% $\pm$ 0.4	84% $\pm$ 0.4	66.0% $\pm$ 0.5
<b>Inv v Uninv</b>	81.4% $\pm$ 0.4	77.6% $\pm$ 0.4	76.8% $\pm$ 0.4	62.8% $\pm$ 0.5
<b>Hinv v Huninv</b>	82.7% $\pm$ 0.4	78.7% $\pm$ 0.4	81.2% $\pm$ 0.4	65.9% $\pm$ 0.5
<b>Pinv v Puninv</b>	88.0% $\pm$ 0.3	85.4% $\pm$ 0.3	82.3% $\pm$ 0.4	74.5% $\pm$ 0.4
<b>Hinv v Pinv</b>	88.1% $\pm$ 0.3	85.6% $\pm$ 0.4	84.3% $\pm$ 0.4	75.8% $\pm$ 0.4
<b>Huninv v Puninv</b>	83.5% $\pm$ 0.4	80.2% $\pm$ 0.4	83.8% $\pm$ 0.4	72.2% $\pm$ 0.5



**Fig 4.10. KDE plots for variables provided.** The top two plots are the distributions from the IKDC And KOOS self-reported assessments while the bottom plots are the distributions from the maximal volitional isometric contraction (MVIC) flexion and extension metrics provided by the Biodex isokinetic dynamometer.

In Figure 4.10, the top two plots are the distributions from the IKDC And KOOS self-reported assessments, which show a clear difference between the cohorts. For the IKDC assessment, the healthy participants have a somewhat bimodal distribution, with the majority having a score near 1 (scaled from 100%), indicating that they have a high level of function and little to no symptoms. The patient reported scores are more spread out, yet there are a fair amount of patients reporting scores near 1 as well. The distribution for the

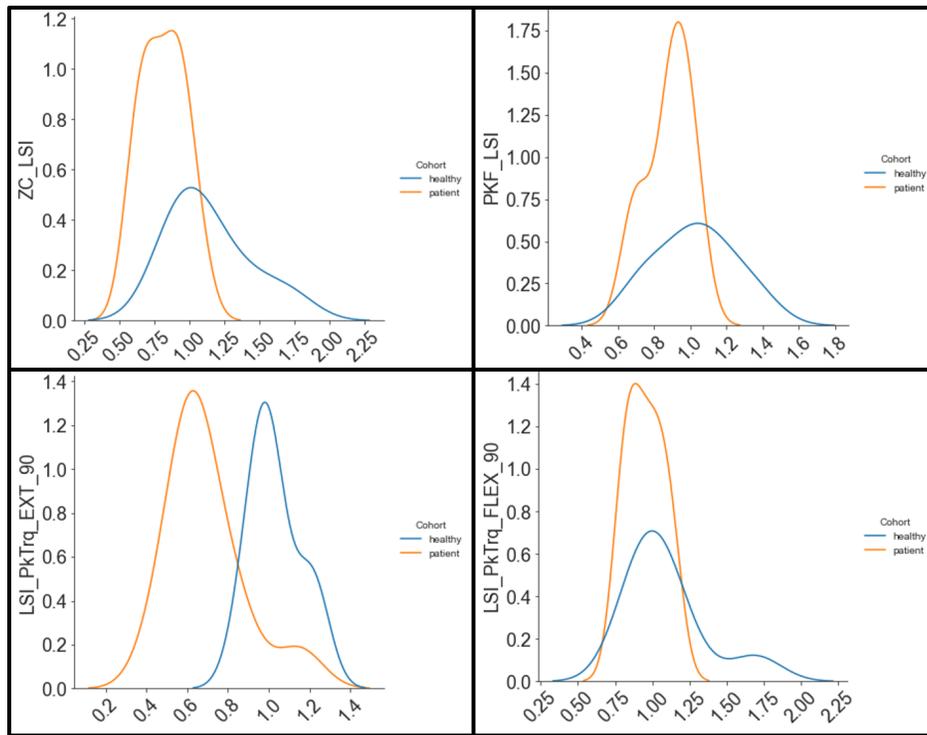
KOOS assessment is almost identical to the IKDC plot, for both the healthy and patient cohorts. The bottom plots are the distributions from the LSI of the maximal volitional isometric contraction (MVIC) flexion and extension measures provided by the Biodex isokinetic dynamometer, which is a measure of muscle strength [85]. The shape of the patient distribution is almost exactly the same between the extension and flexion, yet there is a slightly wider spread for the extension. We do observe a double bump pattern for the patient and healthy cohorts, indicating there are two ranges where each cohort generally scored in. It is interesting to note the mirrored pattern for the healthy cohort between the two measures.



**Fig 4.11. KDE plots of iso 180 features.** Here we see that within the healthy cohort, the median ZC value is very similar between the involved and uninvolved leg. However, there appears to be a clear difference between the median ZC values for the patient cohort.

Next, we look at the iso 180 features and provided LSI measures shown in Figure 4.11. The top two plots are the distributions from features we generated, zero-crossing and peak frequency. Both features show a clear difference between the distributions of the cohorts. The bottom plots are the distributions from the LSI of the peak torque extension and flexion for the iso 180 exercise, provided by the Biodex isokinetic dynamometer. The peak torque

is representative of the dynamic component of muscle performance[86]. We can see the shift in the distributions between the cohorts in the extension plot, but the distributions are very similar for the flexion. Therefore, since the measures provided by the Biodex are similar for both cohorts, they would not be as useful to differentiate between the cohorts, unlike the LSI measures we found for the iso 180 exercise. From the zero-crossing KDE plot, we would be more likely to choose a patient if their ZC-LSI was near a value of 1.



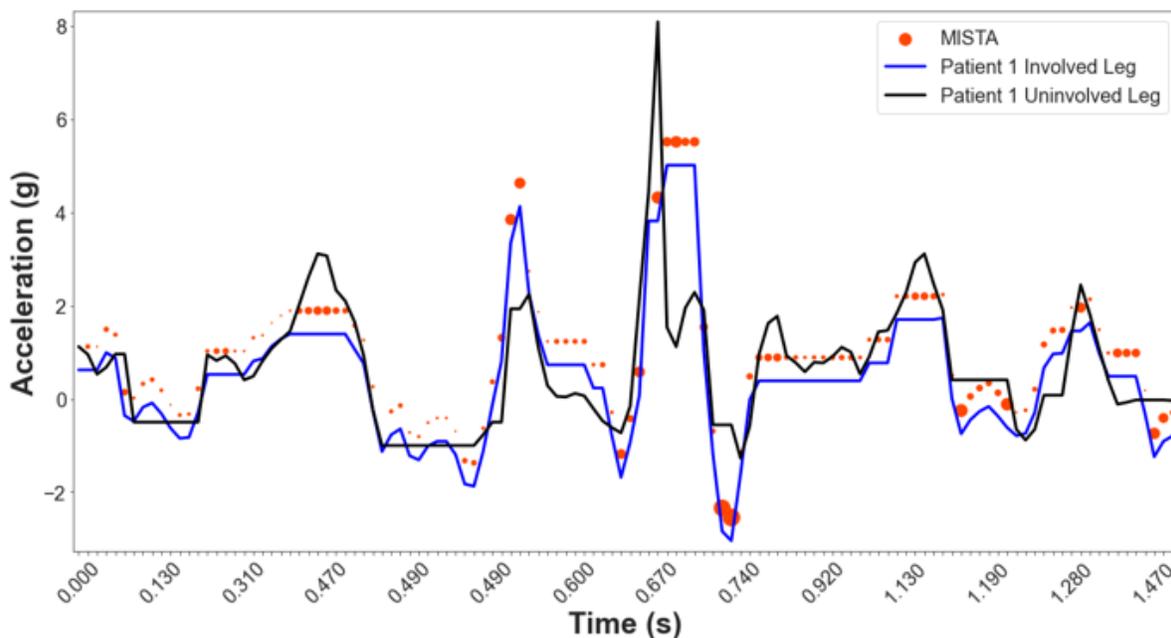
**Fig 4.12. KDE plots of iso 90 features.** Here we see that within the healthy cohort, the median ZC value is very similar between the involved and uninvolved leg. However, there appears to be a clear difference between the median ZC values for the patient cohort.

Then we looked at the features and provided values for the iso 90. In Figure 4.12, we again show the distributions from ZC and peak frequency features. As we previously mentioned for the iso 180 exercise, both features show a clear difference between the distributions of the cohorts. There would be a high probability of selecting a patient if their ZC-LSI value was near 1. The bottom plots are the distributions from the LSI of the peak torque extension and flexion, but for the iso 90 exercise. We observe the shift for the extension measure and the very similar distribution for the flexion measure. Given the results we shown,

the measures provided from the isokinetic dynamometer are similarly distributed, whereas the LSI generated features show vastly different distributions. However, the self-reported assessment scores also separate the cohorts very well.

#### 4.1.5 Applied MISTA Framework

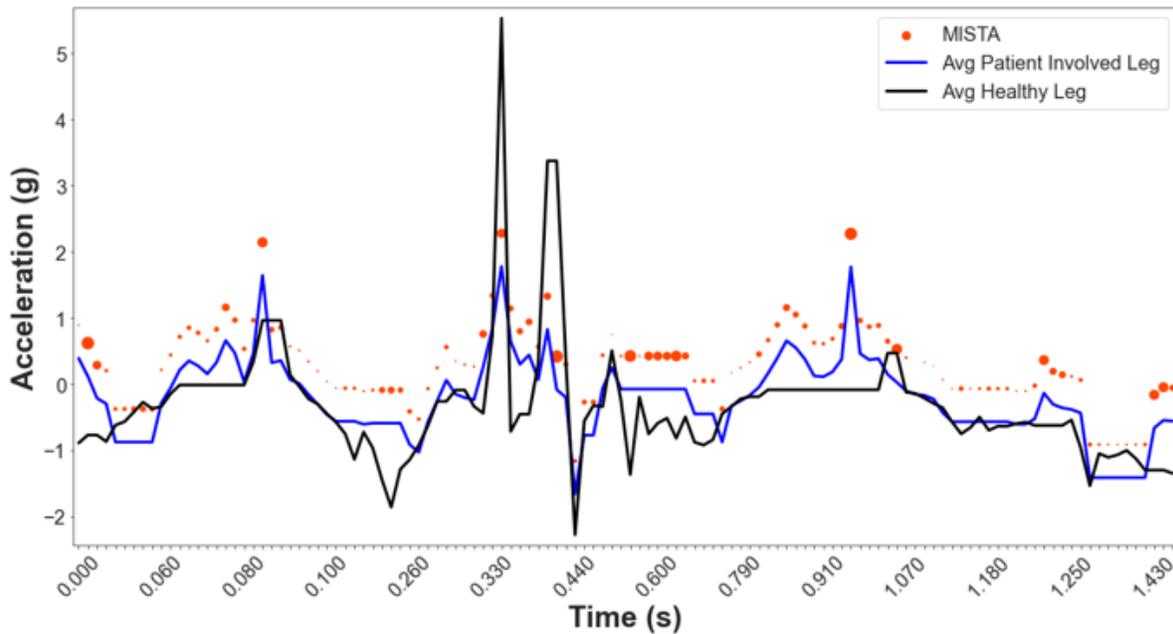
In this section, we demonstrate how MISTA can be used to visualize motion differences across cohorts using the ACLR dataset. However, this time we focus on the gait acceleration data during the treadmill portion of the assessment.



**Fig 4.13.** Comparison between Patient 1’s involved and uninvolved leg.

Figure 4.13 depicts the symmetry between patient 1’s gait cycle. Analyzing the symmetry plot further, we used patient 1’s uninjured leg as the template and their injured leg as the query – patient had undergone ACLR surgery on the right knee. Using the same framework parameters as in Figure 3.3, we notice the uninjured leg is generally higher throughout the gait cycle. Then, there is a clear difference in the downward acceleration from the swing phase, stemming from the patient’s inability to keep their knee elevated. Then, we notice the number of large dots throughout the framework plot, appearing around most of the peaks and valleys. The larger dots indicate statistically significant differences of the max

values during the gait acceleration profile. A key observation occurs just after 1.120 seconds, where the injured leg appears to show no acceleration. This translates to when the patient kept their foot on the ground and their leg straighter for a longer amount of time, followed by a double hitch as they again bring their foot forward. From our observation, Patient 1 had difficulty extending their injured leg and can be detected through our framework plot. Therefore, the patient’s gait currently lacks good symmetry.



**Fig 4.14.** Comparison between the average patient involved leg and the average healthy leg.

Plot B compares the average gait cycles from the injured leg of the 11 patients to the average gait cycle of both legs of the 10 healthy participants. The lower acceleration exhibited by the average patients’ involved leg is consistent with the lower acceleration found in Figure 4.13 by patient 1. Again, we can clearly see the differences, to include the lower peak acceleration exhibited by the average patients involved leg compared to the average healthy participant.

With the MISTA framework, we can also compare a patient’s leg gait cycle to the average gait cycle of the healthy or injured. Figure 4.15 patient 1 vs all healthy participants.

Finally, Figure 4.16 compares patient 1’s motion against healthy participant 1’s motion

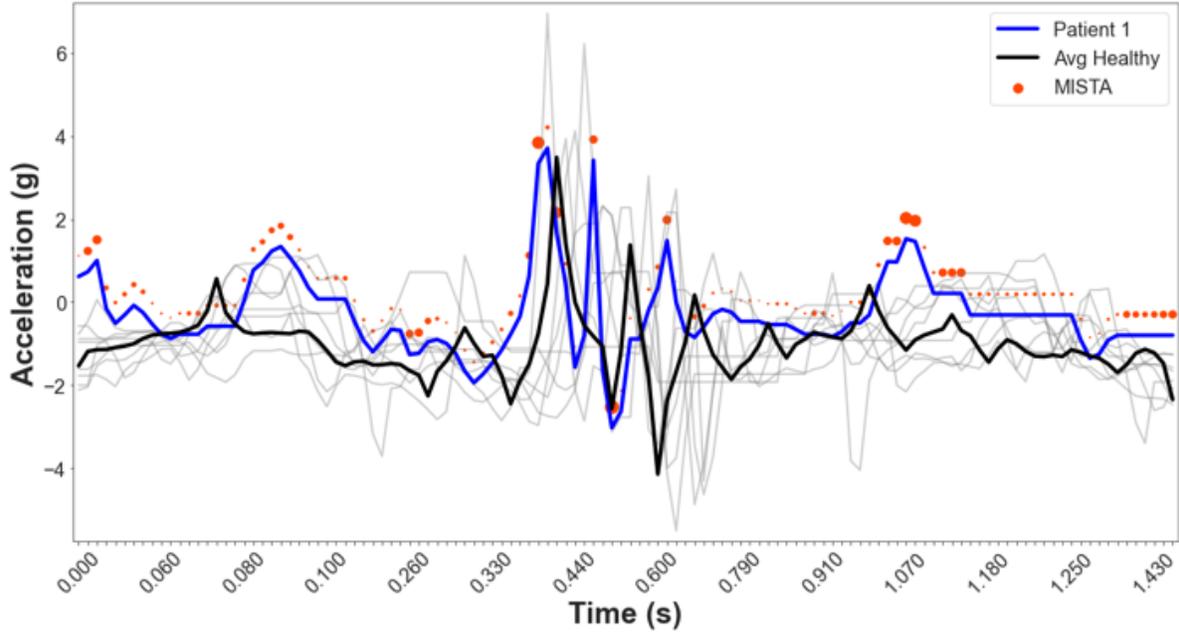


Fig 4.15. Comparison between Patient 1’s involved leg and the average healthy leg. Here we also include all the healthy gait cycles in the background.

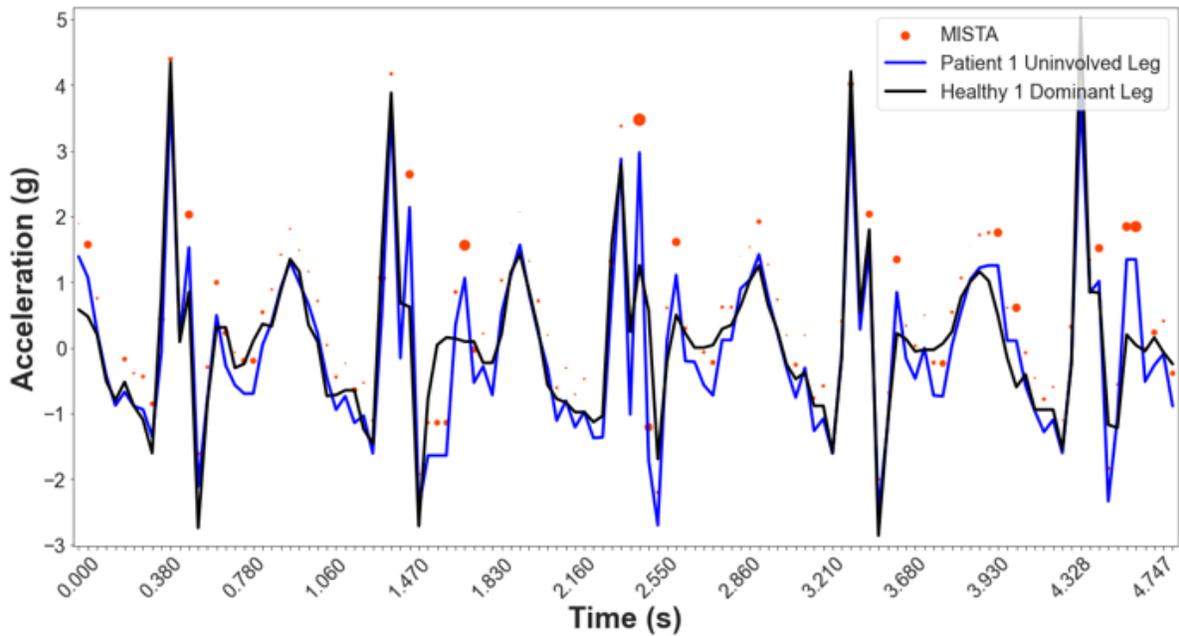
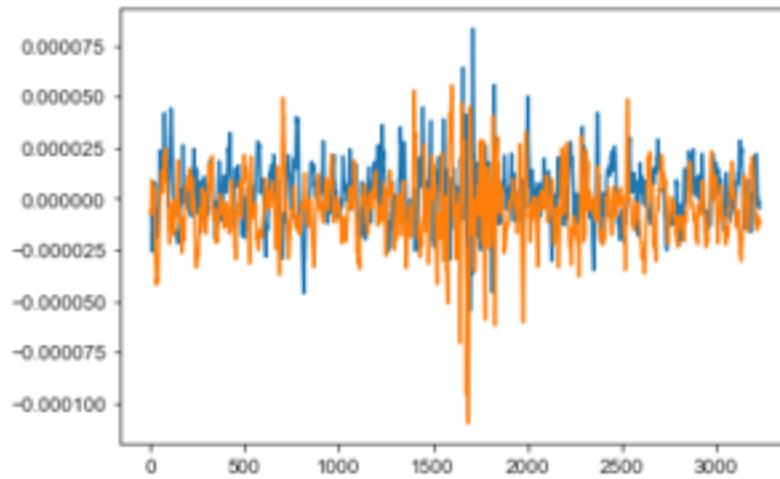


Fig 4.16. Comparison of multiple gait cycles between Patient 1’s uninvolved leg and healthy participant 1’s dominant leg.

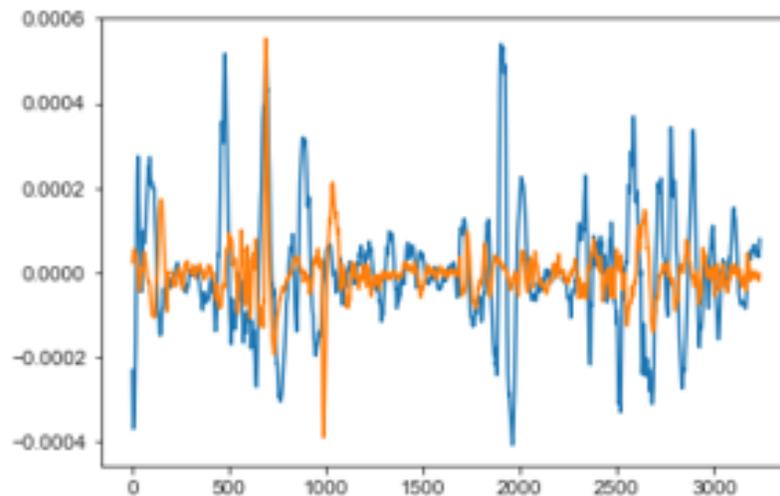
across multiple gait cycles, for the uninvolved leg of patient 1 and the dominant leg of healthy participant 1. Though the plot shows significant differences for the peaks, the overall motion appears to be very similar which is to be expected for the uninjured leg of the patient. So

---

we've discussed four different scenarios that can be visualized, but there are many more possibilities. The comparison can be one gait cycle to another, or across multiple gait cycles.



(a) Healthy 1



(b) Patient 1

**Fig 4.17.** The sEMG data during the same gait cycle for each participant

To validate these results, we observe the sEMG profiles during the same gait cycle we analyzed in our results. In Figure 4.17, the top plot shows healthy participant 1 exhibiting very good symmetry during their gait cycle, while the bottom EMG plot reveals the clear lack of symmetry of patient 1. Not only is the strength of the involved signal (orange) for patient 1 much less throughout the motion, but it is also less consistent. The goal would be to reassess patient 1 in the future to see if their symmetry has improved.

---

#### 4.1.6 Discussion

This work presented a descriptive analysis on data from a cohort of healthy subjects and patients who are recovering from ACLR surgery, using a wearable sensor during two rehabilitation exercises. From these exercises, we were able to detect the individual muscle activations and extract features for comparative analysis. Finally, we leveraged machine learning techniques in an attempt to classify multiple subsets of the data.

Given the six scenarios, we used four classification algorithms to find the best accuracy while treating each muscle activation as an individual sample. From our results, the random forest algorithm provided the best accuracy across all the scenarios. There was a slight improvement in accuracy when combining the data from both exercises for classification, however, we suspect that more data would help further improve the results. The two scenarios with the highest accuracy compared the patients' involved versus uninvolved leg and the healthy involved versus the patient involved leg. This is inline with our expectation that the patient's involved leg should be much different given the recovery time frame post surgery. We also attempted a four-way classification, treating the cohort and leg as their own class (i.e. healthy-inv, healthy-uninv, etc.), and achieved an accuracy of 73% with the RF model for the iso 180 exercise. Overall, the classification algorithms show that the extracted EMG features are sufficient to distinguish between the cohorts.

To reiterate the results presented when looking at the LSI features, we demonstrated that the measures we generated have different distributions between the cohorts, unlike the measures computed by the Biodex isokinetic dynamometer. Future work would include comparing these LSI generated features to the self-reported assessments with a much larger dataset. Note that the intent of our work is not to compete with existing practices, such as the use of the isokinetic dynamometer, but rather to assist in the need for impairment identification. However, descriptive analysis could be used in combination with or as a stand-alone evaluation tool. The analyses presented serve as an initial step toward correlating strength and functional outcomes for improving the RTS decision making process.

---

#### **4.1.7 Limitations**

Though the results are promising, this work has its limitations. The primary limitation is the number of participants. Though the broader study has over 1200 subjects, only those that wore the wearable sensors were included. Additionally, we did not have enough data to conduct a comparison between the sexes, which would have allowed us to understand if there is a difference in recovery timeline, strength, etc. Another limitation is the sensor itself, where the noise from the sEMG may not be filtered completely. Finally, the results presented are the findings from two of the exercises captured in a clinical setting. The participants performed a series of other exercises yet all were done in a controlled environment. The next step would be to analyze more exercises and capture data in a natural setting.

#### **4.1.8 Conclusion**

In this study, we characterized the muscle activations during two rehabilitation exercises and identified the sEMG features that are significantly different between six cohort scenarios. Our analyses confirm our hypotheses, that 1) healthy subjects can be clearly separated from patients; 2) the healthy subjects do not show a significant difference between their involved and uninvolved legs; and 3) a patient's uninvolved leg is dissimilar to their involved leg in muscle strength. Though we expected the interquartile ranges to be closer between the healthy subjects' legs, the mean and median values were close enough to render the subject-to-subject variability moot. These results demonstrate the effectiveness of using wearable sensors to characterize sEMG muscle activations for a set of exercises, paving the path for future use of wearable technology in ACL rehabilitation assessment. Furthermore, our findings may be used to help with tracking progress through post operative recovery by leveraging advanced analytics and technology in making healthcare decisions.

## **4.2 A Study on Quantifying Motion Quality with Wearable Sensors**

We were provided with the opportunity to conduct a study in conjunction with the University of Virginia's (UVA) Children's Hospital. It is through this collaboration that we had the

---

opportunity to validate our framework with collected motion data.

#### 4.2.1 Background and Motivation

Spinal Muscular Atrophy (SMA) and Duchenne Muscular Dystrophy (DMD) are two of many neuromuscular disorders that affect children worldwide. SMA affects approximately 1 in 11,000 babies born each year, with a life expectancy at less than 2 years. DMD affects 1 in 3,500 to 5,000 males, and is the most common fatal, progressive muscle disorder in boys [87]. Though life expectancy has increased from teens to thirties, boys typically lose their ability to walk in late childhood to early teen years, and have limited use of their arms after their early teens. New treatments have been developed in recent years [88], yet the assessments for measuring progress have remained stagnant.

To quantify the severity and limitation of a patient’s motor function loss, researchers and clinicians have developed several motor function assessments, such as the Children’s Hospital of Philadelphia Infant Test of Neuromuscular Disorders (CHOP-INTEND) [48], the Brooke Upper Extremity Scale [50], and the Hammersmith Functional Motor Scale Expanded (HF MSE) [49]. However, the outcome measures of these assessments are based on ordinal, subjective rating scales, that do not account for the quantity and quality of motion. As Yu et al. [12] pointed out, scaled assessments are susceptible to subjective factors and are influenced by a ceiling effect, making it difficult to track longitudinal progress. Clinicians and researchers alike recognize that the reliance on the human eye to facilitate and score a user’s motion is not the most effective approach for detecting small changes in patient progress, which often go unnoticed [13].

Recent advances in wearable sensing and mobile computing have the potential to fill these gaps. Smartwatches and other wearable devices are embedded with sensors such as accelerometer, gyroscope, magnetometer, etc. that collect a vast amount of data that can be leveraged for digital healthcare. This technology has the ability to measure motion patterns while also capturing micro-movements that have previously gone overlooked. Then by combining these technologies with signal processing and machine learning (ML) algorithms,

---

motor function assessments and rehabilitation may be performed in the comfort of a patient’s home [33]. Previous efforts [89, 6] have attempted to score these motion assessments with various modalities of motion tracking technology, from cameras to inertial measurement units (IMUs). Work such as [90, 61] use cameras and multiple wireless sensors to track upper limb movement to report angles of motion. Physicians have spent countless hours manually evaluating a patient’s performance from video recordings. While this provides the opportunity for a physician to be more thorough in their evaluation, it is still subjected to intra-rater bias. Vision-based activity recognition systems, such as the Microsoft Kinect and Leap Motion systems, use a camera, IR sensor, and IR depth sensor to extract features that can be used for object detection, motion analysis, and other characteristics. While many motion tracking cameras exist in the commercial sector, the vast majority of research for skeletal motion tracking leverages the Microsoft Kinect (v1 and v2) [23, 6, 24, 25, 26]. The Microsoft Kinect SDK provides skeletal joint tracking on a set of 32 joints. The Kinect is able to distinguish semantic features that recognizes based on the premise of a “semantic space” that includes human knowledge about a given activity, object, or attribute feature [91]. Previous and ongoing research are focused on the development of this technology to allow an occupational therapist to quantify the range of motion for upper limbs in three-dimensional space.

These studies focused on classifying the movements according to the assessment scale, rather than quantifying the movement directly which can inject the experimenter’s biases into the algorithm. The scaled assessments also offer little granularity to the incremental changes that may occur with the development of new therapies. Furthermore, motion quality assessment in young children is much more challenging because patients can be nervous or excited during experiments or simply be too young to follow precise instructions [92]. As we demonstrate in our results, this can lead to variance in motion quality scores because of confounding factors such as varying speed of motion, instead of the underlying neuromuscular disorder.

---

The dataset in this study is unlike any previous work on motion quality assessment for a multitude of factors. First, this study had the largest age range of any study in this field, with the inclusion of children as young as 2 years old, and adult patients into their thirties. Second, the range of motor function spans from fully functioning control participants, to patients with DMD and SMA who cannot fully perform the specified motions. This study also includes the extremely rare case of a female DMD patient, 1:50,000,000. This research explores the feasibility of using a minimal set of wearable IMU sensors to develop a score that objectively measures both the quantity and quality of motion. Leveraging time series similarity analysis, the scoring measure will also account for confounding factors like varying motion velocity. Motion quality assessments would benefit from a continuous, quantitative score for three main reasons. First, it would allow these assessments to move away from ordinal based scoring measures, that provide relatively low detail on motion performance and can be affected by the ceiling effect. Second, the score would track most degrees of movement, thereby giving credit for any movement a patient performs. Last, it moves away from the reliance of human observers for assessing performance, which is susceptible to bias.

The purpose of this research is to develop a score that relates both how much motion occurs and how well the motion is performed when performing a movement with wearable sensors. Ultimately through this effort, we address the first research question of proving that ubiquitous sensors can be used to quantify the quality of motion. After describing our methodology, we will introduce a novel study, where we apply our designed score to the data for validation.

#### **4.2.2 Methodology**

Building upon the foundation of curve matching as discussed in section 2.3.2, we present our measure called MeND, or the Mean Normalized Distance, to quantify the motion quality for a subject by comparing their motion trajectory to a reference trajectory. This reference trajectory can either be a standalone baseline or a set of trajectories if obtained from a control cohort. By comparing the motion trajectory, we can directly quantify the quantity

and quality of motion for tracking incremental changes that otherwise go unreported in traditional motor function assessments. The MeND score generation process can be visually represented as a diagram in Figure 4.18 and as pseudo-code in Algorithm 1.

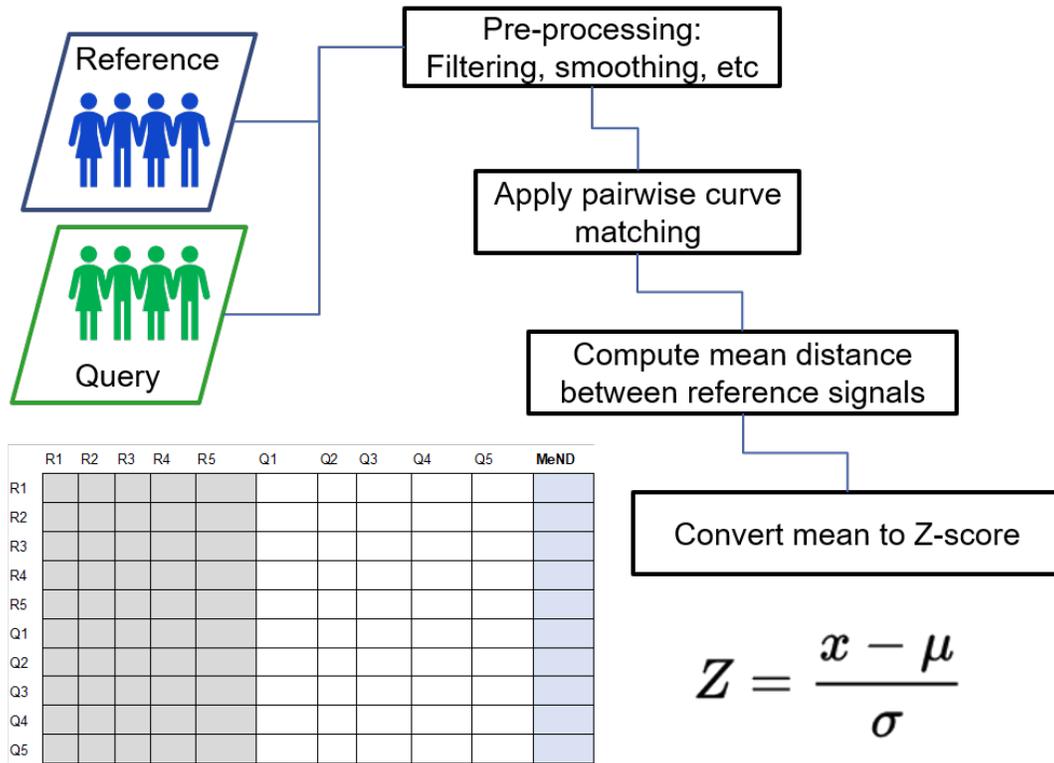


Fig 4.18. Visual representation of generating MeND scores

Given a set of query signals  $V$  and a set of reference signals  $H$ , we must first determine the axis of motion we are attempting to score. For wearable IMU data, a motion can either be evaluated by using the angular velocity derived from the gyroscope or by looking at the acceleration. While there is value in using the acceleration profile or the velocity derived from the integration of the acceleration, we found that a motion can be assessed by its angular velocity along the direction of the motion. Then, we aggregate the signals into a single list  $S$  and apply pre-processing techniques, such as a 3rd order Butterworth lowpass filter.

Next, we compare the two sets of signals with the DTW distance metric, just as we described in section 2.3.2 of curve matching. Equation 4.2 [43] details the formula for DTW, which states that given two time series  $i$  and  $j$ , the optimal alignment is defined as the

---

**Algorithm 1: Mean Normalized Distance (MeND)**

---

**Input:** $H = \{h_1, \dots, h_m\}$  Reference Motion Signals; $V = \{v_1, \dots, v_n\}$  Query Motion Signals;**Output:**

Similarity score of each signal;

**Procedure:**

- 1: *Determine axis of motion path*
  - 2: **Stack** data into one list  $S_{m+n}$
  - 3: *Pre-processing: Lowpass Butter filter*
  - 4: **For** each signal in  $S$
  - 5:       5th order Butterworth Lowpass filter
  - 6: **end for**
  - 7: Initialize new array  $T_{m+n,m+n}$
  - 8: **For**  $i$  in range( $m+n$ )
  - 9:       **For**  $j$  in range( $m+n$ )
  - 10:            $T =$  Pairwise curve matching with DTW
  - 11:        $dist(v) = \sum_{h \in H} \frac{T(m,n)}{|H|}$  *Average score across reference signals*
  - 12: MeND = Convert  $V$  to Z-score
  - 13: **Return** MeND
- 

minimized cost between pairs of indices, subject to a path that satisfies three conditions: boundary conditions, monotonicity, and continuity.

$$dtw(i, j) = d(i, j) + \min \begin{cases} dtw(i-1, j) \\ dtw(i, j-1) \\ dtw(i-1, j-1) \end{cases} \quad (4.2)$$

where  $d(i, j)$  corresponds to the Euclidean distance between the DOF of frame  $i$  of one motion segment and the DOF of frame  $j$  of the second motion segment. The optimal path is then calculated through matrix  $D(i, j)$  by starting at the last frames of each of the signals and moving backwards through the smallest distance values. The DTW distance is then the sum of the distances along the optimal path, which may also be normalized according to the new time duration found after alignment.

---

Given the calculated distances  $T$ , we developed a motion quality score shown in Equation 4.3. We take the mean of the normalized distances (MeND) across the reference signals  $h \in H$  as the performance score, relating all the motions to the reference baseline set.

$$DTW - MeND_{score}(v) = \sum_{h \in H} \frac{dtw(h, v)}{|H|} \quad (4.3)$$

The final step is to transform the performance scores to z-scores by treating the reference signals as the sample population and filtering out the outliers to compute the sample mean and standard deviation. Outliers were detected if they were outside 1.5x of the interquartile range. The Z-score transformation measures how many standard deviations away the given value is from population mean.

As a comparative measure to the DTW-MeND score, we also computed ED scores based on the set of statistical features mentioned in Table 4.10. Mathematically, we represent this score in Equation 4.4, where  $\phi(h)$  is the vector of statistical features extracted from each trajectory. The ED distances were transformed to z-scores using the same process for DTW-MeND, where the sample mean and standard deviation were computed from the set of reference signals.

$$ED_{score}(v) = \sum_{h \in H} \frac{\|(\phi(h), \phi(v))\|_2}{|H|} \quad (4.4)$$

Through the process outlined above, we have established a quantitative scoring measure that can now be applied to real world motion data.

#### 4.2.2.1 Subjects

Data was collected from 41 participants from the Pediatric Neuromuscular Clinic at the UVA Children’s Hospital. Inclusion criteria for patients were that they are currently diagnosed with SMA or DMD. Control participants were recruited based on enrolled patients, matching both sex and age ( $\pm 1$  year). See Table 4.7 for the cohort breakdown. Due to the following reasons, some patients’ data has been excluded in the presented results:

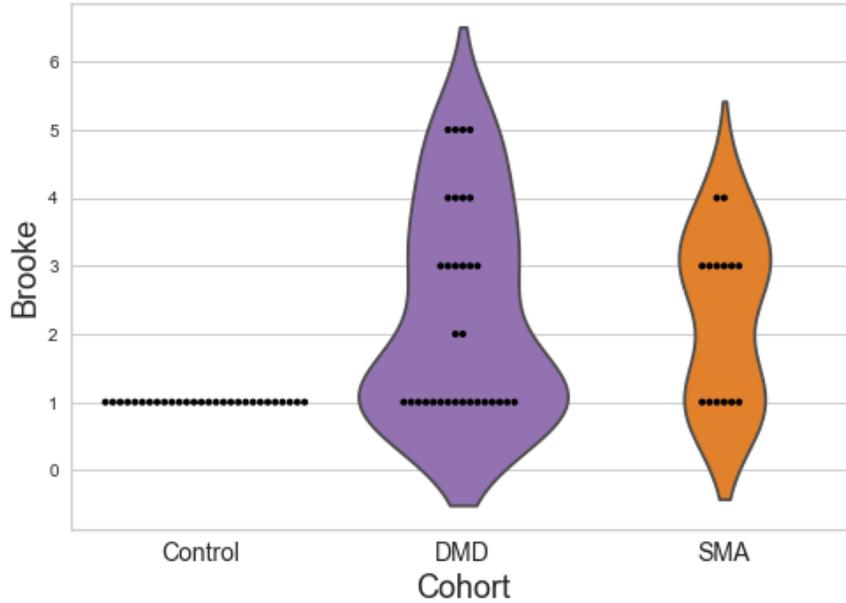
- 
1. For two patients, the sensors stopped recording midway through.
  2. Two of the younger children’s data was indistinguishable for separate activities.
  3. One patient withdrew from the study.

**Table 4.7.** Cohort Profile

<b>Cohort</b>	<b># of Participants</b>	<b>Age Range</b>	<b>Sex</b>
Control	13	2 - 35	8 M, 5 F
SMA	7	2 - 19	6 F
DMD	16	4 - 35	15 M, 1 F

Since the two neuromuscular disorders do not use the same motor function assessments, we evaluated all patients using the Brooke Upper Extremity Scale to provide one standardized metric for comparison [2]. All control participants received a score of 1, indicating full range of motion. See Figure 4.19 for the Brooke scores breakdown of each cohort. Due to the degenerative nature of these disorders, obtaining data from this size of a cohort along with the wide range of age variability is extremely rare. Furthermore, we did not exclude patients who were unable to perform the motions.

Table 4.8 shows the most recent assessment scores for a few patients, with the CHOP-INTEND for kids with SMA and the Brooke Upper Extremity Scale for kids with DMD. We also report how their score has changed in the last 2 years if available. Since the CHOP-INTEND involves more than upper extremity movements, we present the 3 items most related to our motions; however, the scores do not provide any indication of the quality of movement, nor do the reported changes indicate if they affected the specific items listed. For the Brooke scale, 2 of the 3 patients with DMD reported no change in the last 2 years, while the third patient had their score increase by 1, meaning they lost movement. Although the scores for Patients 2 and 12 did not change, this doesn’t mean their functional ability hasn’t changed. With DTW-MeND as a continuous score, our metric will detect smaller, incremental changes to track progress or loss earlier on, which can impact clinical decision



**Fig 4.19.** The Brooke Upper Extremity Scale is a 6-point scale used to classify upper extremity [2]. Higher scores indicate more severe motion impairment. This scale was used as ground truth for all participants.

making for individualized treatment. Motor function assessments should attempt to capture as much concrete data as possible to provide a true evaluation of impairment over time.

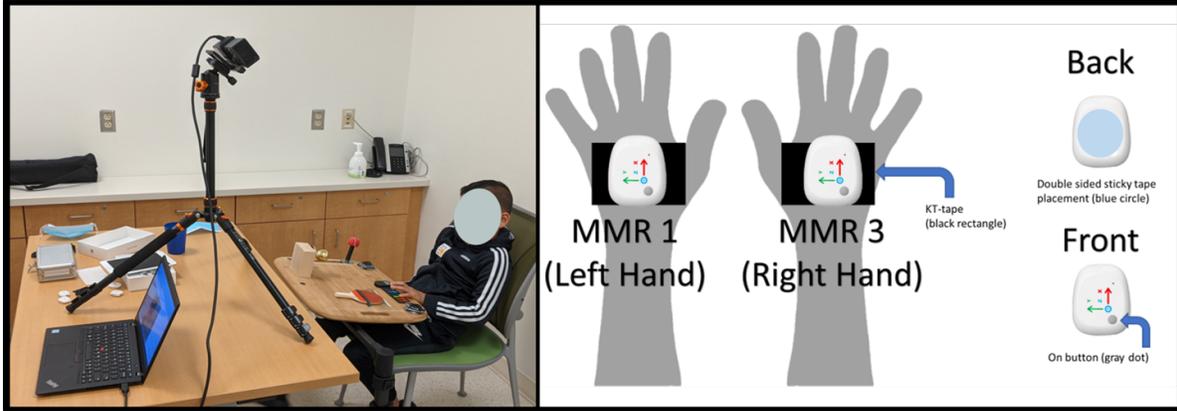
**Table 4.8.** Traditional Motor Function Assessment Scores

ID	Cohort	CHOP	Brooke	+/-
P1	SMA	46	3	CHOP +2 in 2 years
P2	SMA	58	1	CHOP +6 in 1 year
P3	SMA	41	4	CHOP -3 in 2 years
P4	DMD	-	5	No Change
P5	DMD	-	4	No Change
P6	DMD	-	3	Brooke +1 in 2 years

The study was approved by the UVA Internal Review Board (IRB) for Health Sciences Research (HSR), protocol #12161. The participants or the parents of the minor participants provided written informed consent for participation in the study.

#### 4.2.2.2 Experimental Protocol

Figure 4.20 illustrates the full experimental design, with participants sitting in a chair, if able, with a customized, interactive sensory board resting on a reading stand. One MetaMotionR+



**Fig 4.20.** Pilot study set up with two MbientLab MMR+ sensors placed on a participant’s hands and the Microsoft Kinect v2 on a tripod to record and track movements. A custom built sensory board was designed to test specific upper extremity motions related to activities of daily living (ADLs).

(MbientLab, San Francisco, CA, USA) sensor was placed on the top of each of their hand, attached by double sided adhesive, with athletic wrap placed on the skin for additional protection and adhesion. The Microsoft Kinect v2 camera, shown in Figure 4.20, was used to record the movements and indicate the start/stop time for each repetition. Using an iOS app (MetaBase, mbientlab, United States), the accelerometer and gyroscope sensors recorded data at 200 Hz and transmitted via Bluetooth to an iPad Pro (Apple Inc., Cupertino, CA, USA). A high sampling frequency was chosen to aid in the detection of micromovements, due to the limited motor function abilities of some of the patients. The full set of activities for this study are shown in Table 4.9. For this analysis, we will focus on two activities, arm curl (flexion/extension) and forearm rotation (pronation/supination). The participants were asked to perform an arm curl from the starting position of their palm facing up and forearm resting on their leg. After the curl motion, each participant was handed a small paddle to aid in the pronation/supination motion, from a starting position of their forearm resting on a cushioned block with their palm facing up. Each activity was performed two times.

**Table 4.9.** List of activities performed by each participant.

Turn Door Knob	Finger-to-Thumb Test	Cup to Mouth	Shoulder Rotation Sideways
Closed Fist Knock	Open / Close Fist	Bicep Curl	Shoulder Rotation Forwards
Paddle Pronation / Supination	Piano Key Press	Hand Clap	Shoulder Rotation Backwards

---

### 4.2.2.3 Data Processing and Feature Extraction

The first data processing step was to sync the timestamp of the sensor motion data with the video data. This was a necessary step to be able to label the data by each activity. The activities were manually annotated by playing back the data collection session, and marking the start and stop time of each activity. Through this process, we identified three patients who could not be included in the analysis, due to a sensor reading error and indistinguishable motion for the particular activities. Following these procedures, the activities were separated into their own datasets and subjected to signal pre-processing techniques. Due to the noise and drift associated with IMU sensors, the data was filtered using a 3rd order, low-pass Butterworth filter with a cutoff frequency of 0.1 Hz. From the filtered data, trajectory vectors were developed using features extracted from the time and frequency domains. See Table 4.10 for the complete list of features and their derivations. Then, the filtered time series were normalized for further analysis.

### 4.2.3 Results

#### 4.2.3.1 Mitigation of Confounding Factor: Motion Speed Variation

Collecting quality data from children would prove to be extremely difficult for a variety of factors. First, within this cohort, many of the patients either have severely limited upper extremity movement. Second, some of the younger patients either had a very short attention span or had lack of coordination to perform precise movements. Along the same lines, the participants may perform the motions at an unpredictable velocity, which could stem from the emotional difficulty of the clinical environment, travel related factors, or external circumstances [93]. Therefore, speed was inherently a confounding factor. To illustrate the confounding effect of speed, we depict the arm curl motion of healthy participant 2 at three speeds in Figure 4.21: normal (blue), fast (orange), and slow (green). This shows that for a given motion, the speed at which it is performed can vary dramatically which then influences the derived statistical features.

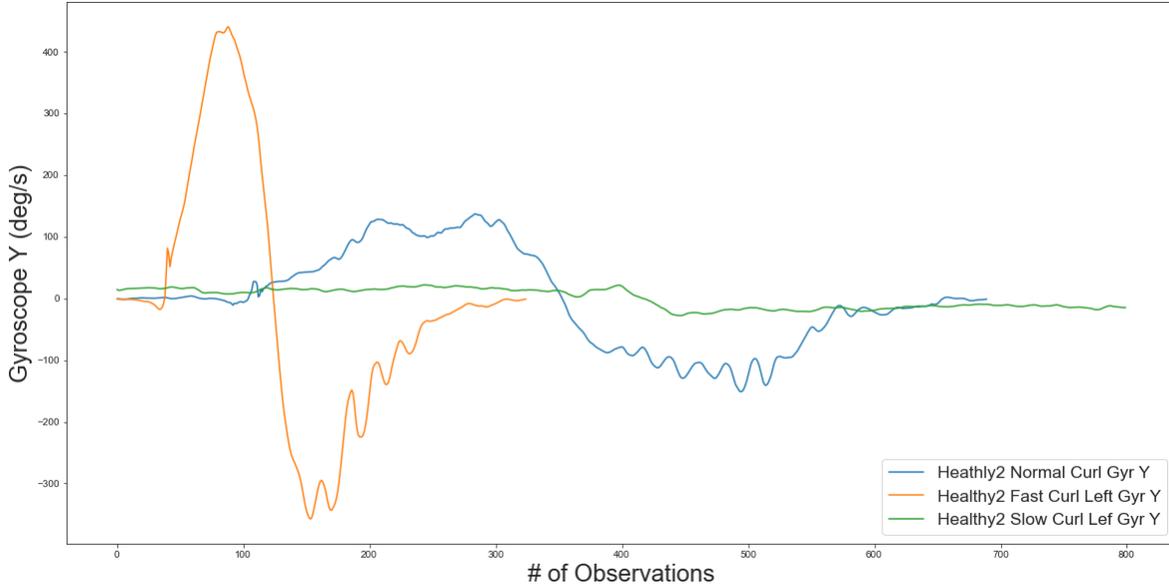
**Table 4.10.** Extracted features

Feature	Formula
Mean	$\bar{x} = \frac{\sum x_i}{N}$
Standard deviation	$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{N}}$
Maximum	$x_{max} = \max\{x(i)\}$
Minimum	$x_{min} = \min\{x(i)\}$
Kurtosis	$kurt = \frac{\sum (x_i - \bar{x})^4}{N\sigma^4}$
Peak to Peak value	$ptop = x_{max} - x_{min}$
Sample Entropy	$sent = -\log\left(\frac{r}{\sqrt{\pi}}\right) - \frac{\Delta^2 \varepsilon (1 - \varepsilon)}{2\zeta_b^2}$
Power Spectral Density	$PSD = S_{xx}(f)$ $= \int_{-\infty}^{\infty} R_{xx}(\tau) e^{-i2\pi f\tau} d\tau = \hat{R}_{xx}(f)$

To assess whether speed is truly a confounding factor, we computed the average ED and DTW-MeND scores for the arm curl across the variable speeds of each arm from two healthy participants. Table 4.11 displays the average scores for each speed. The DTW-MeND scores are very similar in value compared to the large variation shown in the ED scores. This validates our hypothesis that speed is a confounding factor, especially when comparing statistical features, but can be mitigated through the use of DTW.

**Table 4.11.** DTW-MeND vs ED scores for 3 angular velocities.

	DTW	ED
<b>Normal</b>	106.17	492.81
<b>Fast</b>	112.82	291.77
<b>Slow</b>	124.07	398.66



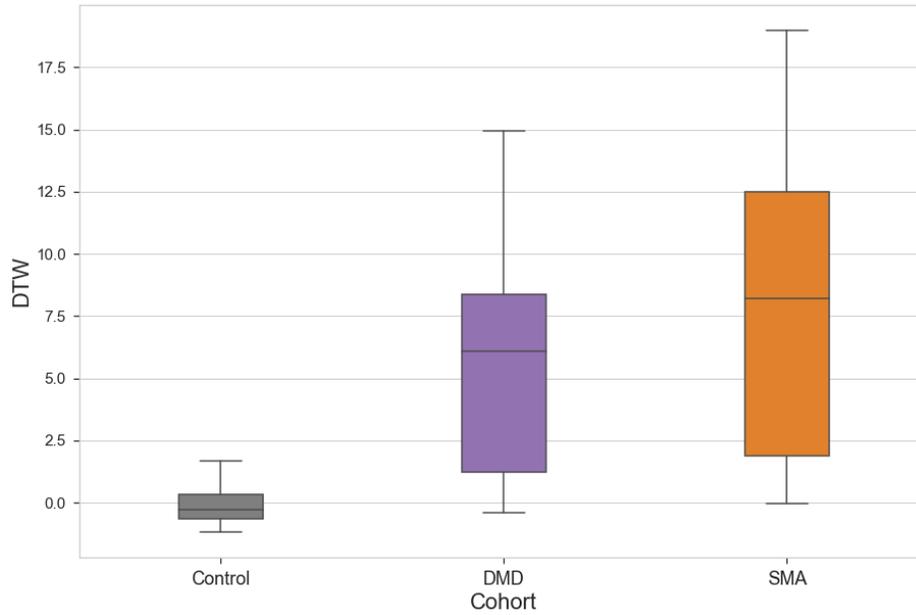
**Fig 4.21.** Arm curl motion performed at 3 different angular velocities (fast/normal/slow).

#### 4.2.3.2 DTW-MeND offers clinically relevant insights

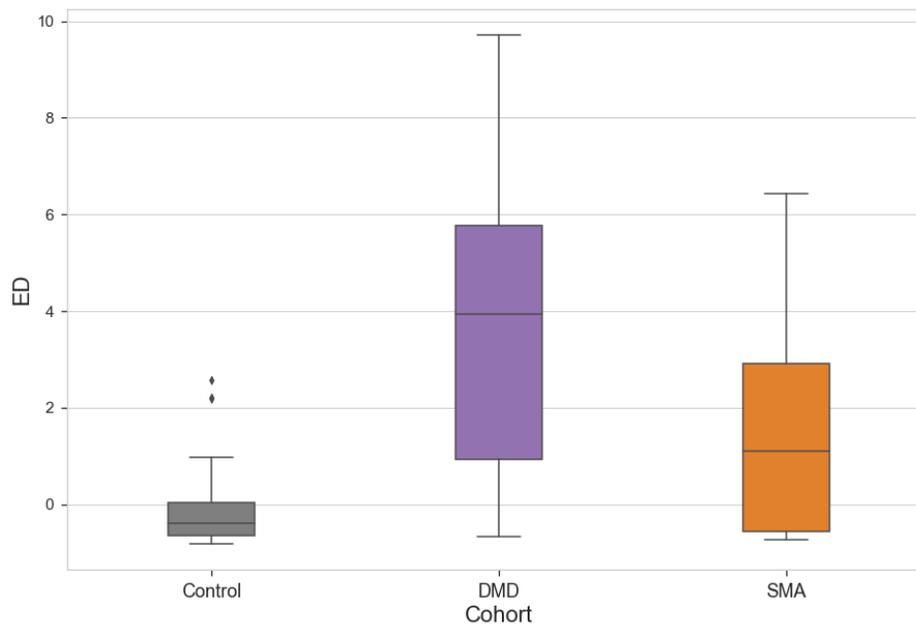
Moving forward, we apply our DTW-MeND score across the arm curl motions and forearm rotations of the entire dataset. Beginning with the arm curl, Figure 4.22(a) depicts the DTW scores for the activity while Figure 4.22(b) illustrates the ED scores. We applied the Mann-Whitney U test [94] with an alpha level of 0.05 to determine if the distributions of each cohort were equal.

The importance of Figure 4.22 underpins one’s understanding of the three cohorts. For the curl motion, there should be a significant difference, as the DTW-MeND boxplot suggests, between the control cohort and the two neuromuscular cohorts. From an observational standpoint, there were patients in both cohorts that could not get their arms into the proper starting position, nor could they complete the full motion. However, some of the patients could perform similar to the control participants. Although the ED boxplot looks like there would be a significant difference between the control and SMA cohorts, the p-value for the ED metric stated otherwise. Similarly, Figure 4.23 displays the DTW-MeND and ED cohort scores for the rotation motion.

For the ED scores of the forearm rotation, the lack of significance between the cohorts



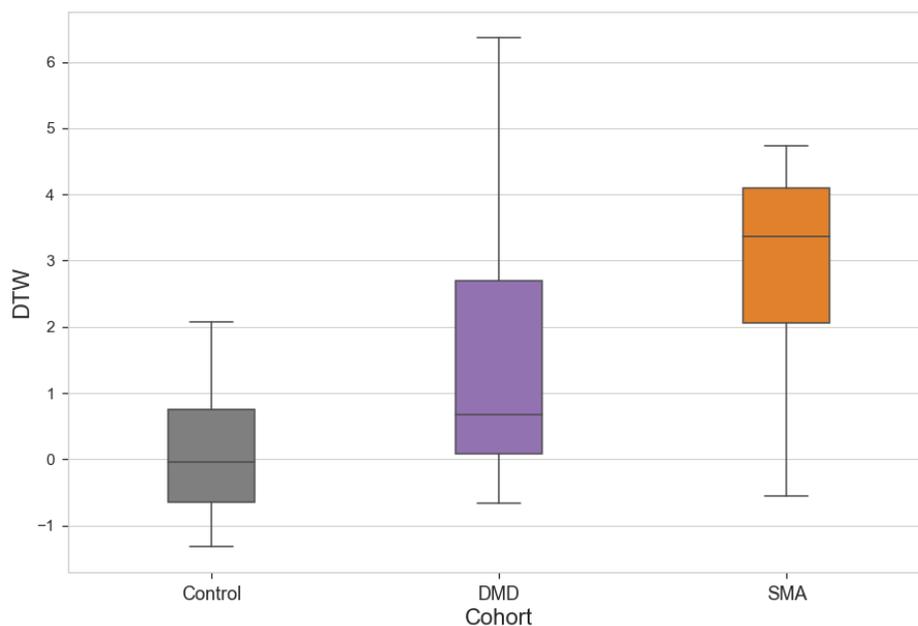
(a) DTW-MeND Boxplot for Arm Curl



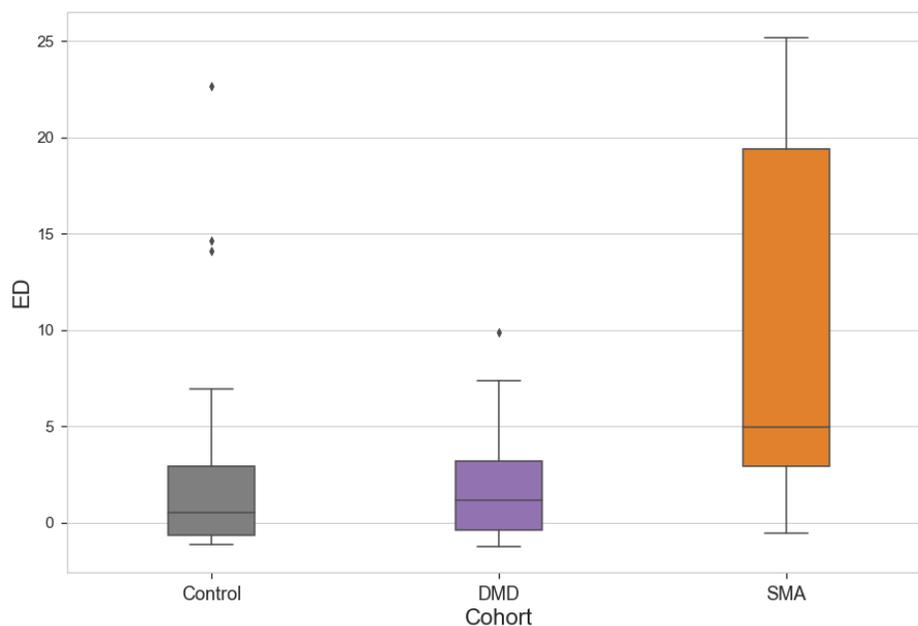
(b) ED Boxplot for Arm Curl

**Fig 4.22.** The top boxplot (a) displays the DTW-MeND curl scores for each cohort. With a p-value  $< 0.001$ , there is a significant difference between the control cohort and the DMD cohort, as well as between the control cohort and the SMA cohort. However, there was no significant difference found between the DMD and SMA cohorts with a p-value = 0.334.

The bottom boxplot (b) depicts the ED curl scores for each cohort. With a p-value  $< 0.001$ , there is a significant difference between the control cohort and the DMD cohort. There was no significant difference found between the DMD and SMA cohorts with a p-value = 0.088, nor was there a significant difference between the control and SMA cohorts with a p-value = 0.095.



(a) DTW-MeND Boxplot for Rotation



(b) ED Boxplot for Rotation

**Fig 4.23.** The top boxplot (a) displays the DTW-MeND rotation scores for each cohort.

A significant difference between the Control and DMD cohorts was determined with a p-value = 0.020, as well as a significant difference between the Control and SMA cohorts with a p-value = 0.037. However, no significant difference was found between DMD and SMA cohorts, with a p-value=0.077. The bottom boxplot (b) depicts the ED rotation scores for each cohort. With p-values > 0.05, there were no significant difference found between any of the cohorts.

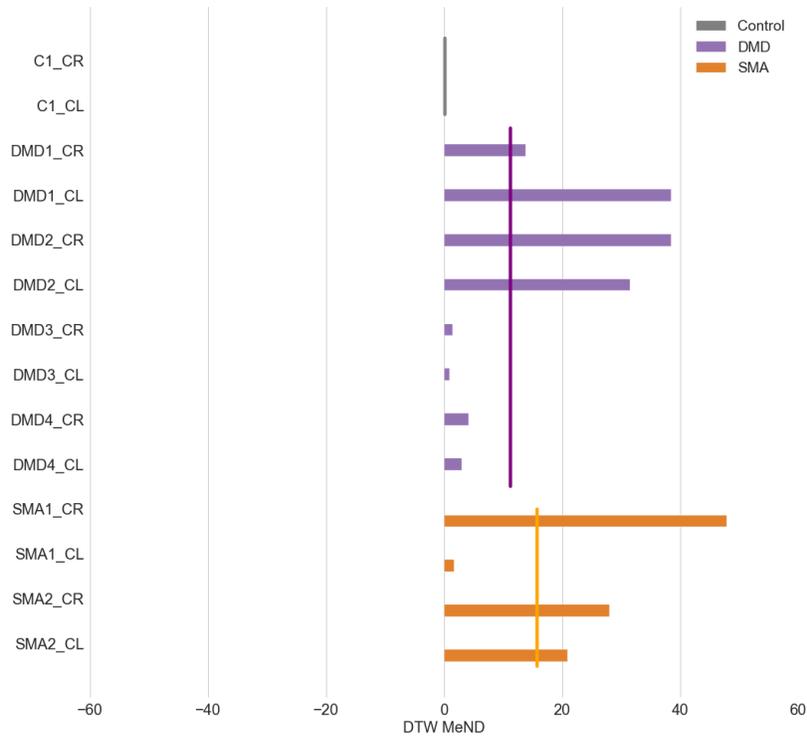
---

does not line up with our observations. It is interesting to see the DMD cohort appear very similar to the control cohort since we anticipated the mean to be higher. However, as previously stated, there are patients who can perform the motion akin to the control cohort. Another observation from this graph is the large variation in the SMA cohort. We suspect this variation may be due to the lack of rotational ability for a few of the patients, which affects the extracted features.

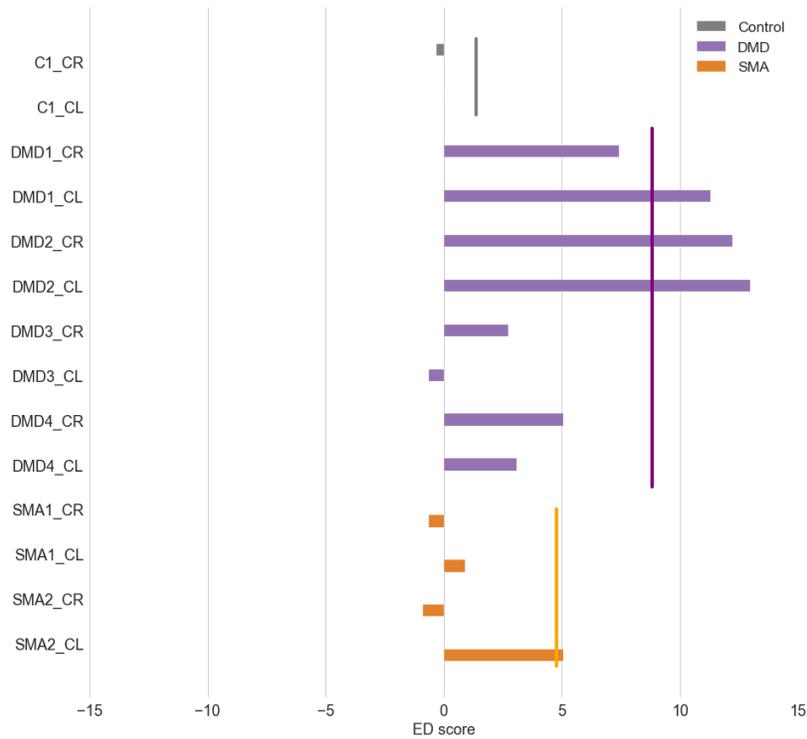
To gain further insight into how individual participants are scored, we observe Figures 4.24 and 4.25 which depict the scores from a small sample for each of the activities relative to the cohort average values.

To draw clinically relevant insights at the individual participant level, we first compare the computed scores to the ground truth video data. In Figure 4.24, we first notice that the ED scores of some patients did not accurately reflect the quality of their motion. For instance, the right arm curl motion of SMA patients 1 and 2 appear to be similar to the motion of the control cohort when using the ED metric. However, the DTW-MeND metric reflects that the motion was actually dissimilar to the control motion, which is inline with the motion seen in the video. Along the same lines, we see the opposite effect happen, where the ED metric for right arm curl motion of DMD patient 3 suggests dissimilarity to the control cohort, despite the motion being very close to that of the control cohort. The DTW-MeND metric again correctly identifies this, by scoring the motion closer to that of the control cohort.

Second, we observed large differences in the DTW-MeND scores between a patient's left and right arms, which were cause for further inspection. For SMA patient 1, the DTW-MeND scores in Figures 4.24 and 4.25 clearly show a stark contrast for the right and left motions. To assess whether this difference is significant, we performed a T-test to compare their left arm versus their right, which was found to be statistically significant with a p-value of 0.0001. Knowing that SMA patient 3 is left arm dominant, both through observation and patient reported data, we further support this difference as an accurate reflection in

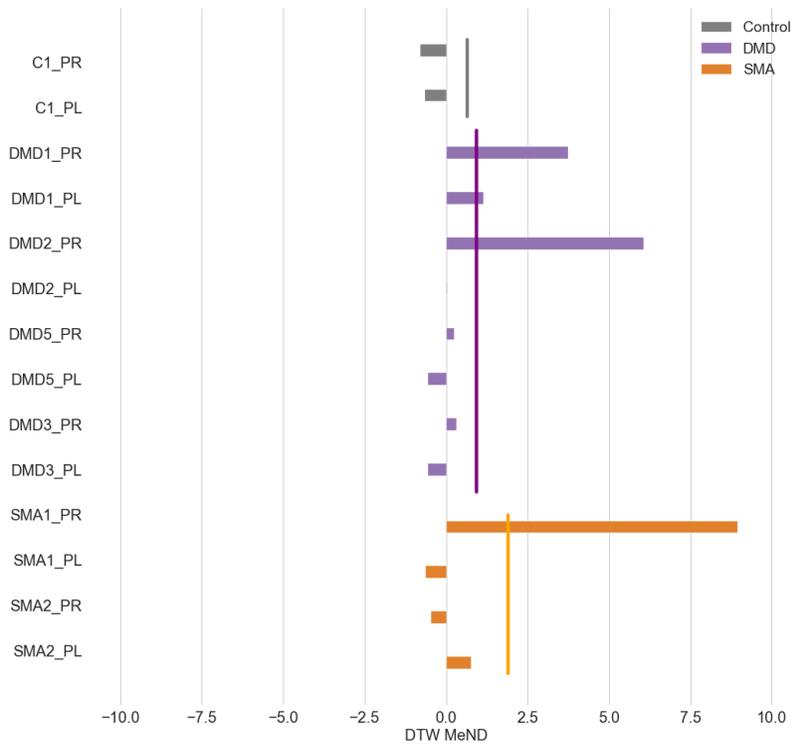


(a) Sample of DTW-MeND scores

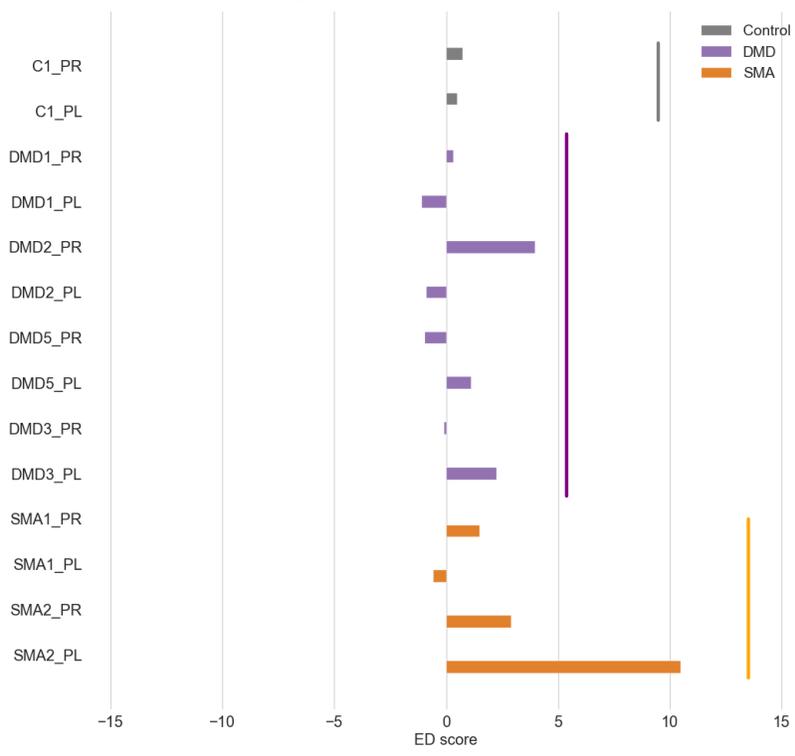


(b) Sample of ED scores

**Fig 4.24.** Sample comparison of DTW-MeND (a) and ED scores (b) for Arm Curl. The vertical lines represent the cohort average values for the specific activity and scoring measure.



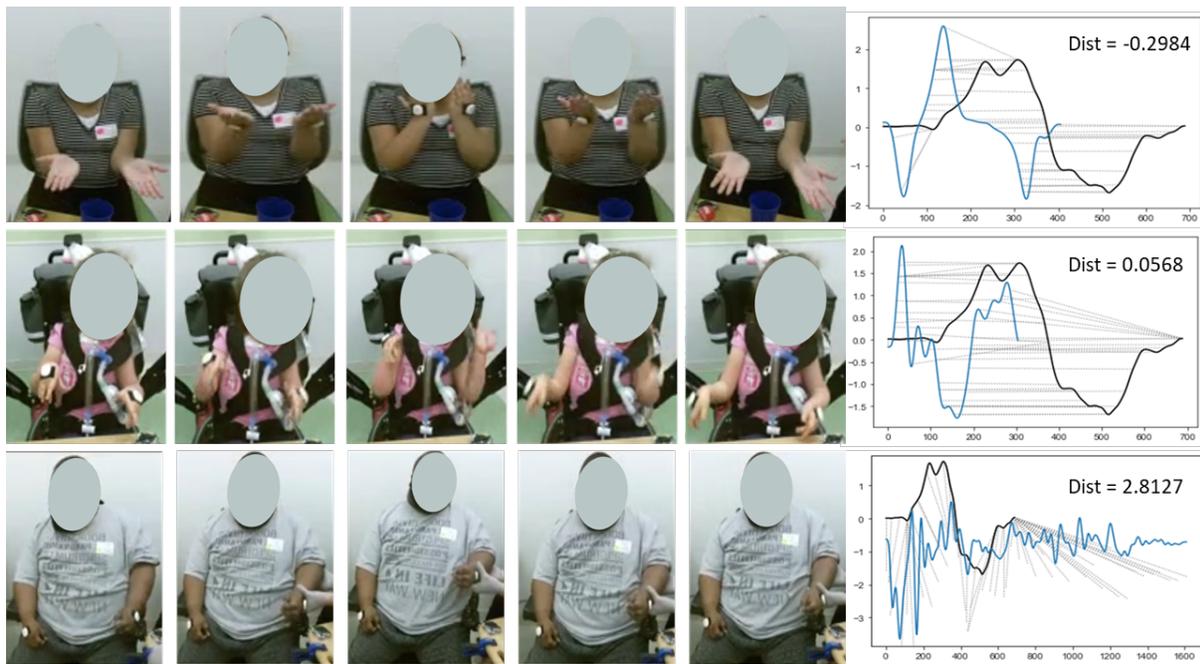
(a) Sample of DTW-MeND scores



(b) Sample of ED scores

**Fig 4.25.** Sample comparison of DTW-MeND (a) and ED scores (b) for Arm Rotation. The vertical lines represent the cohort average values for the specific activity and scoring measure.

the DTW-MeND score. This same analysis can be applied to other patients, such as SMA patient 2, DMD patient 2, and a few others, where their DTW-MeND score for either motion shows distinct differences, and is supported by t-test p-values less than 0.1. These significant differences hint at another insight to be drawn from the DTW-MeND score, which is arm dominance. DMD patient 2 claimed to be right arm dominant yet the DTW-MeND score and video observation suggest otherwise. We found that the difference between their left arm motion and right arm motion was statistically significant, with a p-value of 0.0030. The interpretation of arm dominance through the DTW-MeND score offers a data point that physicians can reference as motor function either progresses or regresses.



**Fig 4.26.** DTW-MeND quantifies the quality of motion subjectively, to indicate how similar the motion is to the control cohort. From top to bottom, the patients' normalized distance score of the arm curl increases as the quality of the curl motion decreases.

The DTW-MeND score applied to this data is inherently a proxy for understanding the severity of impairment. Figure 4.26 displays three patients as they attempt to complete one arm curl. To the right of each patient is their motion trajectory (blue) aligned with the average control reference trajectory (black) and their computed DTW-MeND score in the top right corner. An increasing DTW-MeND score corresponds to progressive loss of functional

---

ability. Patient 27 (top) has a score of -0.2984 and is able to perform an arm curl completely. Patient 15 (middle) has a score of 0.0568, which is reflective of their inability to complete the upper part of the curl and of their inability to rotate their palms face up. Despite the inverted trajectory, our metric still gives some credit to the patient for the motion they could perform. Patient 5 (bottom) has severe motor function loss and could barely lift their arm up, which is echoed both in their score of 2.8127 and in their trajectory.

### 4.2.3.3 Predictive Modeling of Motion Impairment

Now that the motions have been scored, we then leveraged supervised machine learning algorithms to determine if our DTW-MeND score can identify the different cohorts better than the ED score. Classifying DMD and SMA patients can be beneficial to objectively detect differences in movement patterns between neuromuscular disorders, which can then be used to provide higher quality feedback to doctors and occupational therapists. As [80] et al noted, some of the top machine learning classifiers include random forest (RF), linear discriminant analysis (LDA), support vector machine (SVM), k-nearest neighbor (KNN), and logistic regression. Due to the limited size of the dataset, it is recommended to use leave-one-out-cross-validation to aid in assessing the performance of each the classification models.

A RF classification model fits a number of decision trees with different sub-samples of the dataset. Then, the predictions are decided by taking the average of the output of the decision trees. LDA is an algorithm that focuses on dimensionality reduction by transforming the data into a set of component axes that maximize the variance of the data while also maximizing the separation between multiple classes. A SVM is a supervised machine learning algorithm that uses linear decision hyperplanes (or boundaries) to classify points in the feature space, where each hyperplane attempts to separate the data based on their labels. KNN is another supervised learning algorithm designed for classification that aims to predict the associated class by selecting the 'K' nearest points according to a specified distance metric.

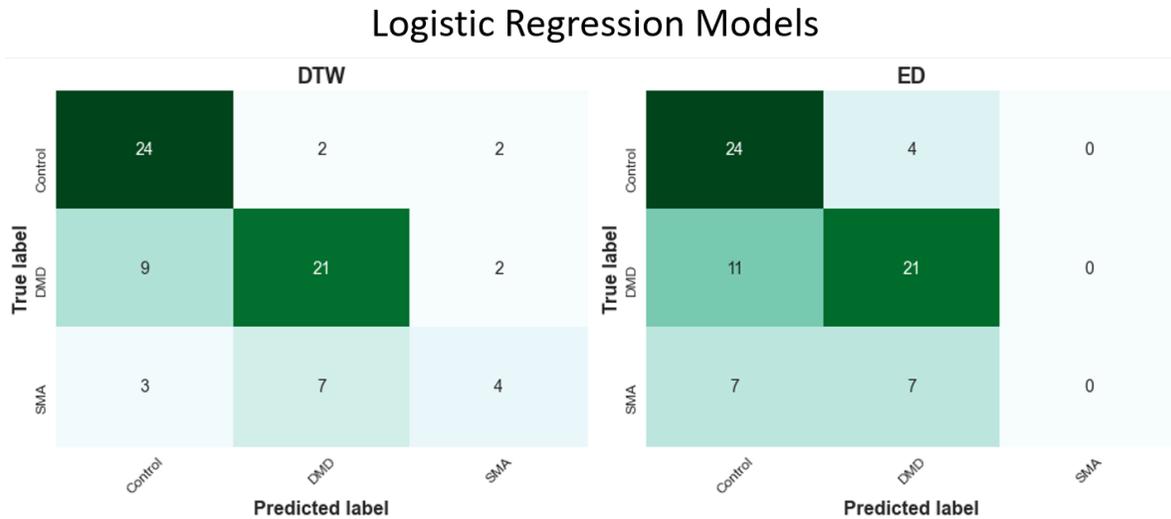
For the arm curl motion, the classification accuracy for each algorithm is shown in Table

4.12 below. In 3 out of the 5 algorithms, our DTW-MeND score performed better than the ED score, while the other 2 algorithms performed the same.

**Table 4.12.** Classification Accuracy for Arm Curl

	<b>RF</b>	<b>LDA</b>	<b>KNN</b>	<b>SVM</b>	<b>LogR</b>
<b>DTW-MeND</b>	62%	58%	59%	61%	66%
<b>ED</b>	45%	59%	54%	61%	61%

Another useful tool to measure the performance of the classification models is the confusion matrix. Confusion matrices offer insight into how each algorithm classified the samples and which classes were predicted most correct. Two scores that can be generated from the confusion matrix are precision and recall. Precision refers to the ratio of true positives over the total of true positives and false positives. Recall refers to the ratio of true positives over the total of true positives and false negatives. Since the logistic regression model achieved the highest classification accuracy for the arm curl motion, we take a closer look at the confusion matrices in Figures 4.27 and 4.28.



**Fig 4.27.** Confusion matrices for the arm curl Logistic Regression models. DTW-MeND on the left and ED on the right.

The results of the logistic regression confusion matrix in Figure 4.27 show that when using the DTW-MeND scores, the DMD cohort had the highest precision at 70% while the lowest precision was the SMA cohort at 50%. The control cohort had the highest recall at 85%

while SMA cohort had the lowest at 28.5%. For the logistic regression ED confusion matrix, the DMD cohort had the highest precision at 65% but the model was unable to correctly predict the SMA cohort. Similarly, the cohort with the highest recall was the control at 85%.

The confusion matrices for the arm curl motion make it clear that more data is needed to better predict the SMA cohort for this exercise. However, one insight we can gather from these matrices is that most false positives / false negatives appear in the middle left box, where the models predicted control but the true label was DMD. These misclassifications are acceptable since the patients predicted as control were the ones who exhibited greater motor function.

Next, we take a look at the forearm rotation motion, where Table 4.13 displays the classification accuracy for each algorithm.

**Table 4.13.** Classification Accuracy for Arm Rotation

	RF	LDA	KNN	SVM	LogR
<b>DTW-MeND</b>	39%	57%	57%	52%	57%
<b>ED</b>	46%	50%	46%	46%	50%



**Fig 4.28.** Confusion matrices for the arm rotation Logistic Regression models. DTW-MeND on the left and ED on the right.

For the forearm rotation, the logistic regression model again achieved the highest accuracy and thus the associated confusion matrices are shown in Figure 4.28. With the DTW-MeND

---

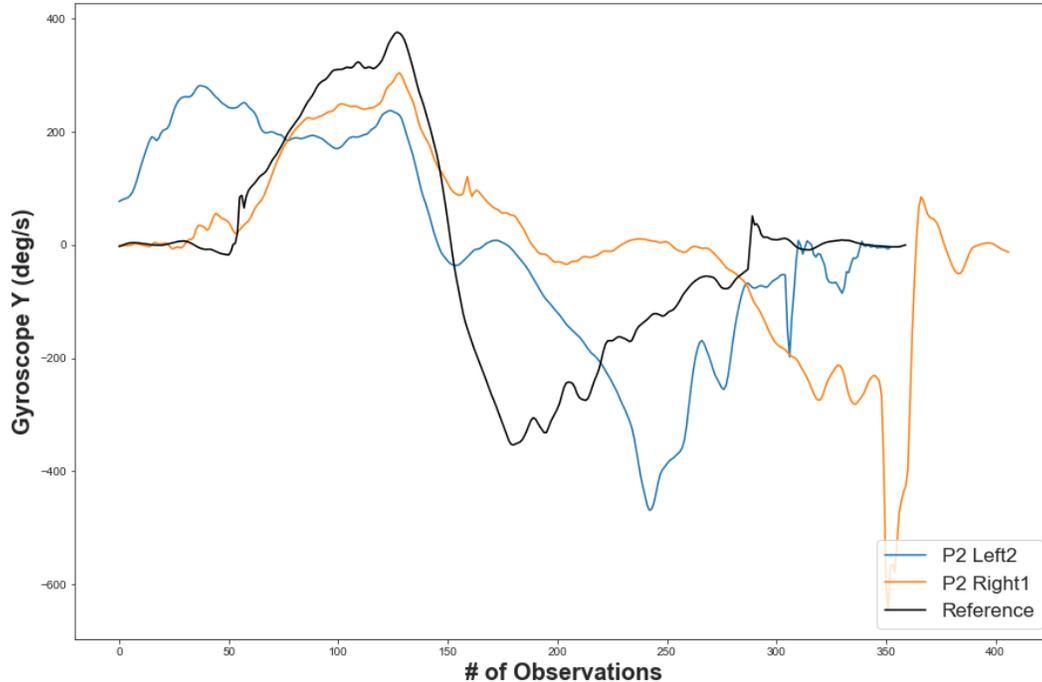
score, the control and SMA cohorts had a precision of 60%, while the DMD precision lagged behind at 46%. The recall rates were not as close, with the control cohort at 77%, SMA at 50% and DMD at 33%. The ED based model predicted most samples as control, causing the control cohort to have the highest precision and recall rates of 54% and 95%. The Logistic Regression ED model did not correctly predict the SMA cohort at all.

While the initial accuracy results of these models may appear to be worrisome, further analysis reveals that the accuracy should be justifiably higher for two reasons. First, the model misclassifying patients as control participants is an accurate assessment, given that those particular patients could perform close to or as good as the control cohort. For those DMD patients predicted as control, they still retain strong motor function ability and most have yet to show any decline in their ability. However, the SMA patients classified as control are those that have received gene therapy and are exhibiting increasing motor function ability at each visit. Secondly, the classification model shown here is a 3-class problem, which resulted in DMD patients being predicted as SMA and vice versa. By converting this to a 2-class problem, the accuracy increases to almost 80%.

#### **4.2.4 MISTA Framework Application**

We continue to validate our framework by looking at the arm curl motion of a patient with SMA from the UEXTEND study. This patient is left arm dominant and has less motor function ability with her right arm. Specifically, she cannot bring her right arm straight up, nor can she pronate her arm into the proper curl starting position. Figure 4.29 depicts the raw arm curl signals of the patient, with the left shown in blue and the right shown in orange, as well as the reference arm curl signal in black. From this chart alone, we can immediately see the difference in the performance between her two arms. The first difference is in the overall lengths of the motion trajectory, which means that the patient took more time to perform the right arm curl than the left. The second observed difference is between the troughs of the motion. The right arm trough is much lower than it is for the left arm.

We then applied our MISTA framework to the patient’s left and right arm individually,

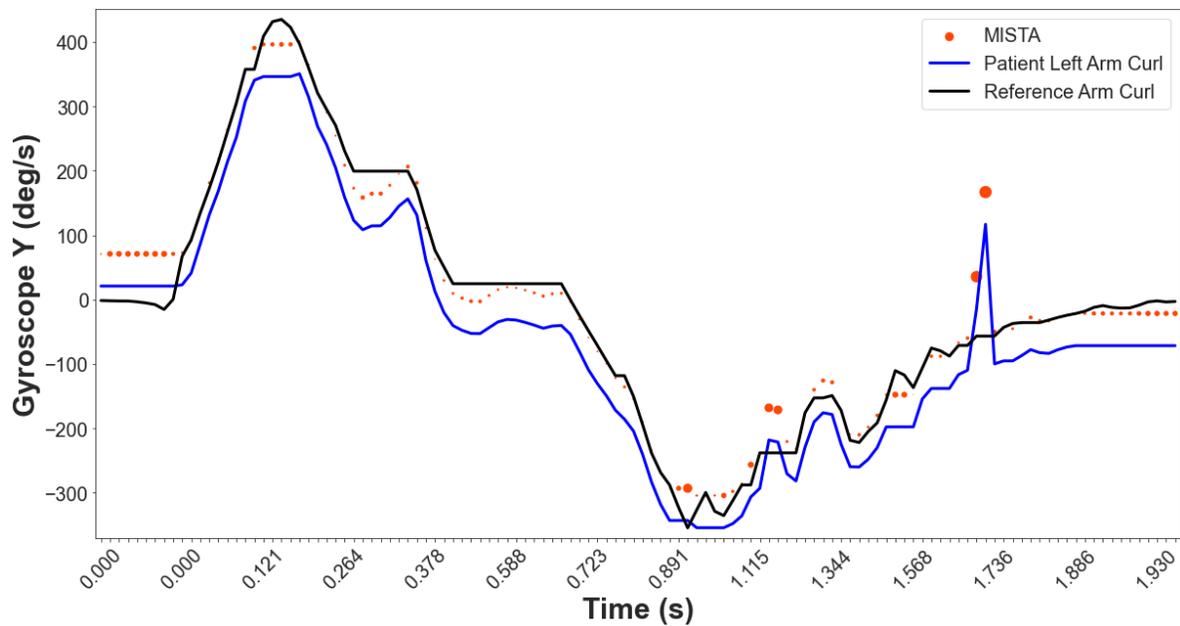


**Fig 4.29.** Raw arm curl motions performed by Patient 2, with blue as their left arm and orange as their right arm. The black line is the average healthy arm curl.

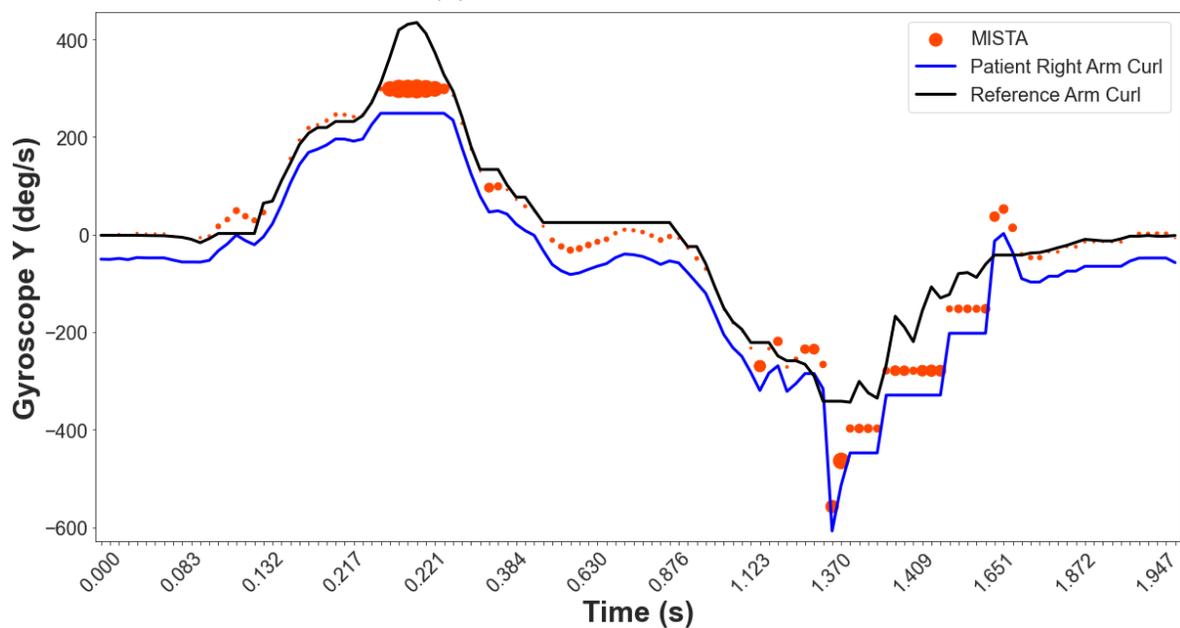
as illustrated in Figure 4.30. The patient’s left arm curl (a) was performed very similar to the reference trajectory, with the exception of a peak around 1.7 seconds. However, we still interpret this as excellent motion quality for this movement. Conversely, the patient’s right arm curl (b) was not performed with the same motion quality, which is noted by the vast amount of large red dots throughout the trajectory. Since we have the exact times that the motion deviated from the reference trajectory, we look to the ground truth video data to understand the cause for the deviation. For the deviation occurring at the top peak, just after 0.217, this is due to the patient swinging her arm inward, towards her chest, as she attempted to complete the arm curl. After speaking with the clinician leading this study, this observed behavior is a compensation technique commonly found in patients with weakened motor function ability.

#### 4.2.5 Future Work

Future work will expand the number of motions assessed and the number of participants, combined with statistical significance tests to provide more quantitative results. Our ap-



(a) Patient 2 left arm curl



(b) Patient 2 right arm curl

**Fig 4.30.** The MISTA framework applied to Patient 2's arm curl motions, with the left arm on top and the right arm on bottom.

---

proach will also extend to a longitudinal study, by comparing a patient’s score to their own baseline for a motor function loss/gain evaluation. This measure will allow us to track the minimal clinically important difference (MCID) over time to assess progress with novel treatments. From a clinical standpoint, a doctor could track their patients movements in reference to the control population, as well as to a patient’s initial baseline. Furthermore, extracting speed invariant features can be useful for other use cases beyond neuromuscular disorder function assessments.

#### **4.2.6 Conclusion**

We built upon the foundation of DTW to design a scoring measure that quantifies the quality of motion using signals generated from wearable sensors. We demonstrated that the DTW-MeND score can effectively mitigate speed as a confounding factor by comparing a motion performed at three different speeds. We validated our measure by applying it to a novel upper extremity motor function assessment. This is the first study we are aware of that look at patients from two neuromuscular disorder cohorts and captures upper extremity motion related to activities of daily living using wearable sensors. This study has the largest age range we’ve found in these communities, with the least restrictive eligibility criteria, and also includes the extremely rare case of female patient with DMD. Our analysis included clinically relevant insights that can aid in tracking patient progression for micro-level motions. The results also demonstrated the ability of DTW-MeND scores to reflect the severity of impairment and infer arm dominance. Finally, we showed how machine learning algorithms can use the DTW-MeND score to differentiate between the cohorts.

This impact of this work has the potential to revolutionize the way clinical trials are performed moving forward by recording objective measures of motion rather than subjective, scaled based measures. The intent is not to replace nor to discredit any of the current motor function assessments, but rather to suggest a new method of scoring those assessments with mobile sensing technology on a continuous scale. The authors in [53] recognized that there is little consensus in the assessment motions and outcomes for upper extremity interventions,

---

where many tools only focus on whether a patient can or cannot complete a task. By tracking incremental changes, this study will also help determine which treatments are, or are not, supporting progression or sustainment of muscle structure and function, and align effective therapies appropriately.

## Chapter 5

### Discussion

The primary contribution of this research is the proposal of a novel measure for quantifying the quality of motion. We provided evidence that curve matching was suitable for mitigating confounding factors, such as motion speed. We then applied our measure to data gathered from wearable sensors for validation. To our best knowledge, this is the first work that applies DTW to upper extremity motor function data collected from wearable sensors. We are also the first work to develop a combined study on patients with Spinal Muscular Atrophy (SMA) and Duchenne Muscular Dystrophy (DMD), spanning the largest age range and with least restricting inclusion criteria. This study also included the extremely rare case of a female patient with DMD. Moreover, this study was the first to attempt to provide quantitative data for upper extremity motor function assessments in these neuromuscular disorder cohorts.

Secondly, we leveraged signal feature analysis techniques to detect and assess rehabilitation progress for patients recovering from ACL reconstruction (ACLR) surgery. Through descriptive analysis, we used extracted features to interpret the differences between the patients and healthy controls. We also used these features in an effort to classify the participants into their respective cohorts, to better understand the threshold where patient performance mirrors healthy control performance. Then, we produced our own set of limb symmetry index (LSI) features and correlated them with the current LSI values, as alternative measures to track rehabilitation progression.

---

Lastly, we proposed a novel framework for modeling deviations in motion trajectory, MISTA, to visualize when those deviations are occurring. We then applied this framework to gait data collected during the ACLR study, where we observed the differences in a gait cycle from each participant, and as a collective cohort. Similarly, the MISTA framework was applied to patient data derived from the neuromuscular disorder study, which confirmed the patient’s arm dominance and provided insight into how patient’s compensate for lack of motor function ability. Overall, our proposed framework advances the state-of-the-art for modeling trajectory analysis.

## 5.1 Clinical Relevance

As we alluded to throughout this research, our proposed methodologies and analysis demonstrate clinical relevance. Primarily, the application of our scoring measure to the neuromuscular disorder motor function assessment study provides an understanding on impairment severity, where we can now track micro-level movements that had previously gone unnoticed. We recognized a critical absence of tools and systems to quantify and objectively capture meaningful motor function in an inclusive way in this population. Our work provides the potential for revolutionizing the way clinical trials are performed moving forward, which will help accelerate the pipeline of new treatments for neuromuscular diseases. Additionally, current clinical trials are not inclusive to patients with cognitive challenges due to the current clinical outcomes requiring strict patient participation. This work is especially important for pediatrics because standard metrics used for adults do not always translate well to children, and vice versa. By tracking incremental changes, this study will also aid in the development of future methodologies to identify which treatments are, or are not, supporting progression or sustainment of muscle structure and function. The proposed methodology could drive the development of software and applications to aid occupational and physical therapy and provide a more objective, quantitative way to evaluate functional upper extremity motions and the quality of those movements. Additionally, this study will allow us to track the minimal clinically important difference (MCID) over time to assess progress with novel treatments.

---

Further clinically relevant insights were observed during the descriptive analysis of the ACLR study, where we could interpret how the muscle activations in a person’s quadriceps can suggest progression status for rehabilitation. We proved the feasibility of including wearable sensors and their role in ACLR rehabilitation going forward. This is especially apparent considering the insights drawn from our motion modeling framework, MISTA. By visually modeling the deviations in motion trajectory, clinicians can better understand range of motion (RoM) limitations, fatigue, and overall quality of motion during a specific exercise. Other insights that can be drawn from DTW-MeND and MISTA are limb dominance and compensation. Limb dominance can be detected by differences in the DTW-MeND score and validated with MISTA plots for each arm. Detecting compensation is not as straightforward, as it requires video confirmation. A physician would need to inspect the timestamps when MISTA had detected significant deviations, and then look at the ground truth video data to see how the patient was moving at that time.

## 5.2 Limitations

The primary limitation of our work is the sample size of our studies. Within the neuromuscular disorder community, our study is considered to be acceptable as data collection is driven by the number of patients available within a certain region. The study design also has its own limitations, with the first being the large range of motor function ability between the healthy controls and the patients. Comparing control functionality proved to be difficult since the patient functionality could be vastly different, yet we were focused on capturing more precision at the lower end of the functional scale and not very concerned about any potential ceiling effects at the range of higher function. Another limiting factor was patient participation, which can be affected by neurodevelopmental challenges, such as autism and intellectual disabilities, resulting in partial or inadequate data collection. This limitation could be mitigated through the inclusion of additional supports to enable improved participation. External factors such as medical illness or additional medical appointments also played a role in limiting data collection attendance.

---

Another limitation of our work is from the sensors themselves. First, there is the accumulation of error from sensor drift that can lead to inaccurate results. Second, is our choice of sensor, which may not have the most sensitive calibration. The number of sensors used and the sensor placement location are also limiting factors. The use of a single IMU sensor on each limb was a trade-off between comfort for the patient, ease of use, and data availability. Furthermore, our analysis focused on a limited set of the motions we collected, leaving the potential for more clinical insights to be drawn.

### 5.3 Future Works

The neuromuscular disorder study presented in this research is part of a multi-modal study that included the collection of ultrasound images of the bicep and forearm, as well as the elbow torque using a handheld dynamometer. From that data, a muscle score describing the structure and function of will be generated and correlated against the motion score we have computed. With this effort, future work includes the expansion of the number of subjects longitudinally tracked to further validate our novel functional measures, the inclusion of additional raters (clinicians, radiographers, physical therapists, and occupational therapists), the expansion to other broader data collection settings (i.e. at home) and the application of these measures other neuromuscular disorders.

Further work must be done to either develop a new assessment for upper extremity motor function or revise the current assessments so the scores are based on the quantifiable measures of motion, such as ROM. The authors in [53] recognize that there is little consensus in the assessment motions and outcomes for upper extremity interventions, where many tools only focus on whether a patient can or cannot complete a task. Therefore, We would like to see an effort undertaken to develop an all-in-one upper extremity motor function that receives consensus from the leading health practitioners that rely on these assessment tools.

The integration of next generation technologies, such as LiDAR, with IMUs will advance telemedicine further than before, with devices readily available like Apple's new iPad Pro. The implementation of new technology can expanded beyond traditional evaluations as pa-

---

tients have shown significant interest in telerehabilitation via game-based movements. Gamification provides a means to retain the attention span of patients by shifting the motions of an assessment tool to be associated with movements found in the game or application. Studies have shown that gamification not only increase motor scale function, but also increase in muscle memory. Advancements in sensor technology will also make 3D kinematic analysis for trajectory reconstruction a viable option in future endeavors. The 3D visualization of one's motion can show an athlete or patient what they are doing wrong at a specific moment in their performance so that the proper adjustments can be made. Then, frameworks like MISTA can move 2D analysis to 3D analysis, which will provide not only when the motion is deviating, but where in physical space the motion is deviating.

## Chapter 6

# Conclusion

Ubiquitous sensing applications are increasing by the day, as technology becomes smaller, faster, and more readily available. Their ability to capture human motion in real time is not only dependent upon the methodologies used to process the data, but also the measures used to quantify those movements. In considering the current landscape of motor function assessments, pairing quantitative outcomes with measures on the quality of the movements can provide informative data for rehabilitation interventions and monitoring recovery. In this thesis, we present a study design that addresses the limitations of current motor function assessments, where we objectively capture motion performance with wearable sensors. By applying our measure, we then fill the need for assessing micro-level movements which can be used as justification for patients to be eligible for and continue to receive novel therapies. By using the ACLR data collected in our study, we demonstrate how signal feature analysis can be used to differentiate between the cohorts and the value in visualizing deviations in motion trajectory. We also provide a brief discussion on the limitations of our work, future research directions, and the clinical impact of this thesis.

## List of Publications

- [1] Gutierrez, R., Kumar, S., McCrady, A., Scharf, R., Blemker, S., Barnes, L., & Boukhechba, M. “A Dynamic Time Warping Metric for Pediatric Motion Quality Assessment: A Feasibility Study.” *Journal of Healthcare Informatics Research* [Accepted pending revision]
- [2] Gutierrez, R., Hart, J., & Boukhechba, M. “Using Inertial Measurement Units (IMU) and Comparative Trajectory Analysis for Modeling Micro-level Human Motion Dysfunction.” In *Proceedings of the 2022 International Conference on Applied Human Factors and Ergonomics*
- [3] Gutierrez, Robert; Hart, Joe; Boukhechba, Mehdi (2022), Post-ACL reconstruction surgery rehabilitation dataset using isokinetic dynamometer and wearable IMUs, Dryad, Dataset, <https://doi.org/10.5061/dryad.66t1g1k4j>
- [4] Gutierrez, R., McCrady, A., Scharf, R., Blemker, S., Barnes, L., & Boukhechba, M. “Upper Extremity Motor Function Assessment: A Protocol for Quantifying Motion with Wearable Sensors.” *JMIR Research Protocols* 2022 [Under Review]
- [5] Gutierrez, R., Hart, J., Boukhechba, M. “Using wearable sensors to detect and assess isokinetic performance following ACL reconstruction.” *PLOS One*. [Under Review]
- [6] Dong, G., Tang, M., Wang, Z., Gao, J., Guo, S., Cai, L., Gutierrez, R., Campbell, B., Barnes, L.E. and Boukhechba, M., 2022. Graph Neural Networks in IoT: A Survey. arXiv preprint arXiv:2203.15935.

- 
- [7] Gutierrez, R., Kumar, S., McCrady, A., Scharf, R., Blemker, S., Barnes, L., & Boukhechba, M. “U-EXTEND: Quantifying Upper Extremity Motor Function and Muscle Architecture.” BMC Medicine 2022. [In Preparation]
- [8] Kumar, S., Gutierrez, R., Datta, D. “Shape analysis for pediatric upper motion quality assessment.” [Under Review]

## References

- [1] J. Wang, Y. Chen, S. Hao, X. Peng, and L. Hu, “Deep learning for sensor-based activity recognition: A survey,” *Pattern Recognition Letters*, vol. 119, pp. 3–11, 3 2019.
- [2] M. H. Brooke, R. C. Griggs, J. R. Mendell, G. M. Fenichel, J. B. Shumate, and R. J. Pellegrino, “Clinical trial in duchenne dystrophy. I. The design of the protocol,” *Muscle & Nerve*, vol. 4, pp. 186–197, 5 1981.
- [3] S. R. Bethi, A. RajKumar, F. Vulpi, P. Raghavan, and V. Kapila, “Wearable Inertial Sensors for Exergames and Rehabilitation\*,” in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 4579–4582, 2020.
- [4] O. D. Lara and M. A. Labrador, “A Survey on Human Activity Recognition using Wearable Sensors,” *IEEE Communications Surveys & Tutorials*, vol. 15, no. 3, pp. 1192–1209, 2013.
- [5] G. Bleser, B. Taetz, M. Miezal, C. A. Christmann, D. Steffen, and K. Regenspurger, “Development of an Inertial Motion Capture System for Clinical Application,” *i-com*, vol. 16, pp. 113–129, 8 2017.
- [6] S. Lee, Y. Lee, and J. Kim, “Automated Evaluation of Upper-Limb Motor Function Impairment Using Fugl-Meyer Assessment,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, pp. 125–134, 2018.

- 
- [7] M. Boukhechba and L. E. Barnes, "Swear: Sensing using wearables. Generalized human crowdsensing on smartwatches," in *International Conference on Applied Human Factors and Ergonomics*, pp. 510–516, Springer, 2020.
- [8] R. M. Koldenhoven and J. Hertel, "Validation of a Wearable Sensor for Measuring Running Biomechanics," *Digital Biomarkers*, vol. 2, no. 2, pp. 74–78, 2018.
- [9] K. Taylor, R. Portela, and L. Gebbie, "Healthy outlook for wearables as users focus on fitness and well-being," 2 2021.
- [10] E. Preatoni, S. Nodari, and N. F. Lopomo, "Supervised Machine Learning Applied to Wearable Sensor Data Can Accurately Classify Functional Fitness Exercises Within a Continuous Workout," *Frontiers in Bioengineering and Biotechnology*, vol. 8, p. 664, 2020.
- [11] A. L. Olsen, L. H. Magnussen, L. H. Skjaerven, J. Assmus, M.-A. Sundal, R. Ostelo, and L. I. Strand, "Movement quality evaluation and its correlation with recommended functional measures in hip osteoarthritis," *Physiotherapy Research International*, vol. 25, p. e1848, 10 2020.
- [12] L. Yu, D. Xiong, L. Guo, and J. Wang, "A Compressed Sensing-Based Wearable Sensor Network for Quantitative Assessment of Stroke Patients," *Sensors*, vol. 16, no. 2, 2016.
- [13] J. C. Miangolarra, "La biomecánica demuestra "cómo y por qué" se producen lesiones o enfermedades neurológicas," 5 2018.
- [14] H. Bennett, K. Davison, J. Arnold, M. Martin, S. Wood, and K. Norton, "RELIABILITY of a MOVEMENT QUALITY ASSESSMENT TOOL to GUIDE EXERCISE PRESCRIPTION (MOVEMENTSCREEN)," *International journal of sports physical therapy*, vol. 14, pp. 424–435, 6 2019.

- 
- [15] K. Pierzchlewicz, I. Kępa, J. Podogrodzki, and K. Kotulska, “Spinal Muscular Atrophy: The Use of Functional Motor Scales in the Era of Disease-Modifying Treatment,” *Child Neurology Open*, vol. 8, p. 2329048X211008725, 1 2021.
- [16] S. J. Erbaugh, “Role of Visual Feedback in Observational Motor Learning of Primary-Grade Children,” *Perceptual and Motor Skills*, vol. 60, pp. 755–762, 6 1985.
- [17] J. Taborri, J. Keogh, A. Kos, A. Santuz, A. Umek, C. Urbanczyk, E. van der Kruk, and S. Rossi, “Sport Biomechanics Applications Using Inertial, Force, and EMG Sensors: A Literature Overview,” *Applied bionics and biomechanics*, vol. 2020, p. 2041549, 6 2020.
- [18] Z. Lu, X. Chen, Q. Li, X. Zhang, and P. Zhou, “A Hand Gesture Recognition Framework and Wearable Gesture-Based Interaction Prototype for Mobile Devices,” *IEEE Transactions on Human-Machine Systems*, vol. 44, no. 2, pp. 293–299, 2014.
- [19] D. Roetenberg, H. Luinge, and P. Slycke, “Xsens MVN: Full 6DOF Human Motion Tracking Using Miniature Inertial Sensors,” 2009.
- [20] S. Alavi, D. Arsenault, and A. Whitehead, “Quaternion-Based Gesture Recognition Using Wireless Wearable Motion Capture Sensors,” *Sensors (Basel, Switzerland)*, vol. 16, p. 605, 4 2016.
- [21] C. Li, L. Yu, and S. Fei, “Real-Time 3D Motion Tracking and Reconstruction System Using Camera and IMU Sensors,” *IEEE Sensors Journal*, vol. 19, no. 15, pp. 6460–6466, 2019.
- [22] U. Côté-Allard, G. Gagnon-Turcotte, F. Laviolette, and B. Gosselin, “A Low-Cost, Wireless, 3-D-Printed Custom Armband for sEMG Hand Gesture Recognition,” *Sensors (Basel, Switzerland)*, vol. 19, p. 2811, 6 2019.

- 
- [23] R. J. Adams, M. D. Lichter, A. Ellington, M. White, K. Armstead, J. T. Patrie, and P. T. Diamond, "Virtual Activities of Daily Living for Recovery of Upper Extremity Motor Function," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 1, pp. 252–260, 2018.
- [24] A. Scano, M. Caimmi, M. Malosio, and L. M. Tosatti, "Using Kinect for upper-limb functional evaluation in home rehabilitation: A comparison with a 3D stereoscopic passive marker system," in *5th IEEE RAS/EMBS International Conference on Biomedical Robotics and Biomechatronics*, pp. 561–566, 2014.
- [25] B. Dehbandi, A. Barachant, D. Harary, J. D. Long, K. Z. Tsagaris, S. J. Bumanlag, V. He, and D. Putrino, "Using Data From the Microsoft Kinect 2 to Quantify Upper Limb Behavior: A Feasibility Study," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 5, pp. 1386–1392, 2017.
- [26] L. P. Lowes, L. N. Alfano, B. A. Yetter, L. Worthen-Chaudhari, W. Hinchman, J. Savage, P. Samona, K. M. Flanigan, and J. R. Mendell, "Proof of concept of the ability of the kinect to quantify upper extremity function in dystrophinopathy," *PLoS Curr*, vol. 5, 2013.
- [27] R. Haghghi Osgouei, D. Soulsby, and F. Bello, "Rehabilitation Exergames: Use of Motion Sensing and Machine Learning to Quantify Exercise Performance in Healthy Volunteers," *JMIR rehabilitation and assistive technologies*, vol. 7, pp. e17289–e17289, 8 2020.
- [28] A. Vakanski, J. M. Ferguson, and S. Lee, "Metrics for Performance Evaluation of Patient Exercises during Physical Therapy," *International journal of physical medicine & rehabilitation*, vol. 5, p. 403, 6 2017.

- 
- [29] H. Su, S. Liu, B. Zheng, X. Zhou, and K. Zheng, “A survey of trajectory distance measures and performance evaluation,” *The VLDB Journal*, vol. 29, no. 1, pp. 3–32, 2020.
- [30] X. Yu and S. Xiong, “A Dynamic Time Warping Based Algorithm to Evaluate Kinect-Enabled Home-Based Physical Rehabilitation Exercises for Older People,” *Sensors (Basel, Switzerland)*, vol. 19, p. 2882, 6 2019.
- [31] S. Adam, B. Coward, G. DeBerry, C. Glazier, E. Magnusson, and M. Boukhechba, “Investigating Novel Proximity Monitoring Techniques Using Ubiquitous Sensor Technology,” in *2021 Systems and Information Engineering Design Symposium (SIEDS)*, pp. 1–6, 2021.
- [32] M. Rigoni, S. Gill, S. Babazadeh, O. Elsewaisy, H. Gillies, N. Nguyen, P. Pathirana, and R. Page, “Assessment of Shoulder Range of Motion Using a Wireless Inertial Motion Capture Device-A Validation Study,” *Sensors*, vol. 19, p. 1781, 10 2019.
- [33] F. Porciuncula, A. V. Roto, D. Kumar, I. Davis, S. Roy, C. J. Walsh, and L. N. Awad, “Wearable Movement Sensors for Rehabilitation: A Focused Review of Technological and Clinical Advances,” *PM&R*, vol. 10, no. 9, Supplement 2, pp. S220–S232, 2018.
- [34] R. H. Osgouei, D. Soulsbv, and F. Bello, “An Objective Evaluation Method for Rehabilitation Exergames,” in *2018 IEEE Games, Entertainment, Media Conference (GEM)*, pp. 28–34, 2018.
- [35] B. Steenwinckel, D. De Paepe, S. Vanden Hautte, P. Heyvaert, M. Bentefrit, P. Moens, A. Dimou, B. Van Den Bossche, F. De Turck, S. Van Hoecke, and F. Ongenae, “FLAGS: A methodology for adaptive anomaly detection and root cause analysis on sensor data streams by fusing expert knowledge with machine learning,” *Future Generation Computer Systems*, vol. 116, pp. 30–48, 2021.

- 
- [36] A. Melendez-Calderon, C. Shirota, and S. Balasubramanian, “Estimating Movement Smoothness From Inertial Measurement Units,” *Frontiers in bioengineering and biotechnology*, vol. 8, p. 558771, 1 2021.
- [37] L. Liu, X. Chen, Z. Lu, S. Cao, D. Wu, and X. Zhang, “Development of an EMG-ACC-Based Upper Limb Rehabilitation Training System,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 3, pp. 244–253, 2017.
- [38] A. S. Abdull Sukor, A. Zakaria, N. Abdul Rahim, L. Kamarudin, R. Setchi, and H. Nishizaki, “A hybrid approach of knowledge-driven and data-driven reasoning for activity recognition in smart homes,” *Journal of Intelligent & Fuzzy Systems*, vol. 36, pp. 1–12, 1 2019.
- [39] A. Murad and J.-Y. Pyun, “Deep recurrent neural networks for human activity recognition,” *Sensors*, vol. 17, no. 11, p. 2556, 2017.
- [40] T. R. Derrick and J. M. Thomas, “Time Series Analysis: The Cross-Correlation Function,” in *Innovative Analyses of Human Movement* (N. Stergiou, ed.), vol. Chapter 7, pp. 189–205, Champaign, Illinois: Human Kinetics Publishers, 2004.
- [41] C. S. Möller-Levet, F. Klawonn, K.-H. Cho, and O. Wolkenhauer, “Fuzzy clustering of short time-series and unevenly distributed sampling points,” in *International symposium on intelligent data analysis*, pp. 330–340, Springer, 2003.
- [42] D. J. Berndt and J. Clifford, “Using Dynamic Time Warping to Find Patterns in Time Series,” in *Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining*, vol. 10 of *AAAIWS’94*, pp. 359–370, Seattle, WA, USA:, 1994.
- [43] D. F. Silva and G. E. A. P. A. Batista, “Speeding Up All-Pairwise Dynamic Time Warping Matrix Calculation,” in *Proceedings of the 2016 SIAM International Conference on Data Mining*, (Philadelphia, PA), pp. 837–845, Society for Industrial and Applied Mathematics, 6 2016.

- 
- [44] E. J. Keogh and M. J. Pazzani, “Derivative dynamic time warping,” in *Proceedings of the 2001 SIAM international conference on data mining*, pp. 1–11, SIAM, 2001.
- [45] D. Folgado, M. Barandas, R. Matias, R. Martins, M. Carvalho, and H. Gamboa, “Time Alignment Measurement for Time Series,” *Pattern Recognition*, vol. 81, pp. 268–279, 2018.
- [46] L. Wang, Y. Sun, Q. Li, T. Liu, and J. Yi, “IMU-Based Gait Normalcy Index Calculation for Clinical Evaluation of Impaired Gait,” *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 1, pp. 3–12, 2021.
- [47] L. M. Schutte, U. Narayanan, J. L. Stout, P. Selber, J. R. Gage, and M. H. Schwartz, “An index for quantifying deviations from normal gait,” *Gait & Posture*, vol. 11, no. 1, pp. 25–31, 2000.
- [48] A. M. Glanzman, E. Mazzone, M. Main, M. Pelliccioni, J. Wood, K. J. Swoboda, C. Scott, M. Pane, S. Messina, E. Bertini, E. Mercuri, and R. S. Finkel, “The Children’s Hospital of Philadelphia Infant Test of Neuromuscular Disorders (CHOP INTEND): test development and reliability,” *Neuromuscular disorders : NMD*, vol. 20, pp. 155–161, 3 2010.
- [49] D. Ramsey, M. Scoto, A. Mayhew, M. Main, E. S. Mazzone, J. Montes, R. de Sanctis, S. Dunaway Young, R. Salazar, A. M. Glanzman, A. Pasternak, J. Quigley, E. Mirek, T. Duong, R. Gee, M. Civitello, G. Tennekoon, M. Pane, M. C. Pera, K. Bushby, J. Day, B. T. Darras, D. De Vivo, R. Finkel, E. Mercuri, and F. Muntoni, “Revised Hammersmith Scale for spinal muscular atrophy: A SMA specific clinical outcome assessment tool,” *PLoS One*, vol. 12, no. 2, p. e0172346, 2017.
- [50] J. M. Florence, S. Pandya, W. M. King, J. D. Robison, L. C. Signore, M. Wentzell, and M. A. Province, “Clinical trials in duchenne dystrophy: standardization and reliability of evaluation procedures,” *Physical therapy*, vol. 64, no. 1, pp. 41–45, 1984.

- 
- [51] F.-M. Ar, L. Jaasko, I. Leyman, S. Olsson, and S. Steglind, “The post-stroke hemiplegic patient. 1. A method for evaluation of physical performance,” *Scand J Rehabil Med*, vol. 7, no. 1, pp. 13–31, 1975.
- [52] S. Wolf, P. A. Catlin, M. Ellis, A. Archer, B. Morgan, and A. Piacentino, “Assessing Wolf Motor Function Test as Outcome Measure for Research in Patients After Stroke,” *Stroke; a journal of cerebral circulation*, vol. 32, pp. 1635–1639, 8 2001.
- [53] J. R. Davids, L. C. Peace, L. V. Wagner, M. A. Gidewall, D. W. Blackhurst, and W. M. Roberson, “Validation of the Shriners Hospital for Children Upper Extremity Evaluation (SHUEE) for Children with Hemiplegic Cerebral Palsy,” *JBJS*, vol. 88, no. 2, 2006.
- [54] B. Galna, G. Barry, D. Jackson, D. Mhiripiri, P. Olivier, and L. Rochester, “Accuracy of the Microsoft Kinect sensor for measuring movement in people with Parkinson’s disease,” *Gait & Posture*, vol. 39, no. 4, pp. 1062–1068, 2014.
- [55] W. Narongwongwathana, P. Punyabukkana, W. Kitisomprayoonkul, and W. Chonnamutt, “WAAM: Wearable Assessment Arm Motion System,” in *2019 12th Biomedical Engineering International Conference (BMEiCON)*, pp. 1–5, 2019.
- [56] J. Kowalczewski, E. Ravid, and A. Prochazka, “Fully-automated test of upper-extremity function,” in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 7332–7335, 2011.
- [57] E. D. O. Simbaña, P. S.-H. Baeza, A. J. Huete, and C. Balaguer, “Review of Automated Systems for Upper Limbs Functional Assessment in Neurorehabilitation,” *IEEE Access*, vol. 7, pp. 32352–32367, 2019.
- [58] Y. Huang, M. Rofouei, and M. Sarrafzadeh, “Automated Wolf Motor Function Test (WMFT) for Upper Extremities Rehabilitation,” in *2012 Ninth International Conference on Wearable and Implantable Body Sensor Networks*, pp. 91–96, 2012.

- 
- [59] J. J. Han, G. Kurillo, R. T. Abresch, E. De Bie, A. Nicorici, and R. Bajcsy, “Upper extremity 3-dimensional reachable workspace analysis in dystrophinopathy using Kinect,” *Muscle & nerve*, vol. 52, pp. 344–355, 9 2015.
- [60] X. Song, S. Chen, J. Jia, and P. B. Shull, “Cellphone-Based Automated Fugl-Meyer Assessment to Evaluate Upper Extremity Motor Function After Stroke,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 10, pp. 2186–2195, 2019.
- [61] B. Oubre, J. Daneault, H. Jung, K. Whritenour, J. G. V. Miranda, J. Park, T. Ryu, Y. Kim, and S. I. Lee, “Estimating Upper-Limb Impairment Level in Stroke Survivors Using Wearable Inertial Sensors and a Minimally-Burdensome Motor Task,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 3, pp. 601–611, 2020.
- [62] P. L. Hudak, P. C. Amadio, C. Bombardier, D. Beaton, D. Cole, A. Davis, G. Hawker, J. N. Katz, M. Makela, and R. G. Marx, “Development of an upper extremity outcome measure: the DASH (disabilities of the arm, shoulder, and head),” *American journal of industrial medicine*, vol. 29, no. 6, pp. 602–608, 1996.
- [63] E. S. Mazzone, A. Mayhew, J. Montes, D. Ramsey, L. Fanelli, S. D. Young, R. Salazar, R. De Sanctis, A. Pasternak, A. Glanzman, G. Coratti, M. Civitello, N. Forcina, R. Gee, T. Duong, M. Pane, M. Scoto, M. C. Pera, S. Messina, G. Tennekoon, J. W. Day, B. T. Darras, D. C. De Vivo, R. Finkel, F. Muntoni, and E. Mercuri, “Revised upper limb module for spinal muscular atrophy: Development of a new module,” *Muscle Nerve*, vol. 55, no. 6, pp. 869–874, 2017.
- [64] R. Mooney, G. Corley, A. Godfrey, L. R. Quinlan, and G. ÓLaighin, “Inertial sensor technology for elite swimming performance analysis: A systematic review,” *Sensors*, vol. 16, no. 1, p. 18, 2016.

- 
- [65] J. J. Irrgang, A. F. Anderson, A. L. Boland, C. D. Harner, M. Kurosaka, P. Neyret, J. C. Richmond, and K. D. Shelborne, “Development and Validation of the International Knee Documentation Committee Subjective Knee Form,” *The American Journal of Sports Medicine*, vol. 29, pp. 600–613, 9 2001.
- [66] N. J. Greco, A. F. Anderson, B. J. Mann, B. J. Cole, J. Farr, C. W. Nissen, and J. J. Irrgang, “Responsiveness of the International Knee Documentation Committee Subjective Knee Form in Comparison to the Western Ontario and McMaster Universities Osteoarthritis Index, Modified Cincinnati Knee Rating System, and Short Form 36 in Patients with Focal Articular Cartilage Defects,” *The American Journal of Sports Medicine*, vol. 38, pp. 891–902, 12 2009.
- [67] E. M. Roos, H. P. Roos, L. S. Lohmander, C. Ekdahl, and B. D. Beynnon, “Knee Injury and Osteoarthritis Outcome Score (KOOS)—Development of a Self-Administered Outcome Measure,” *Journal of Orthopaedic & Sports Physical Therapy*, vol. 28, pp. 88–96, 8 1998.
- [68] M. Salavati, B. Akhbari, F. Mohammadi, M. Mazaheri, and M. Khorrami, “Knee injury and Osteoarthritis Outcome Score (KOOS); reliability and validity in competitive athletes after anterior cruciate ligament reconstruction,” *Osteoarthritis and Cartilage*, vol. 19, no. 4, pp. 406–410, 2011.
- [69] C. Auepanwiriyaikul, S. Waibel, J. Songa, P. Bentley, and A. A. Faisal, “Accuracy and Acceptability of Wearable Motion Tracking for Inpatient Monitoring Using Smartwatches,” *Sensors*, vol. 20, p. 7313, 12 2020.
- [70] I. Bortone, D. Leonardis, N. Mastronicola, A. Crecchi, L. Bonfiglio, C. Procopio, M. Solazzi, and A. Frisoli, “Wearable Haptics and Immersive Virtual Reality Rehabilitation Training in Children With Neuromotor Impairments,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 7, pp. 1469–1478, 2018.

- 
- [71] J. Agel, E. A. Arendt, and B. Bershadsky, “Anterior Cruciate Ligament Injury in National Collegiate Athletic Association Basketball and Soccer: A 13-Year Review,” *The American Journal of Sports Medicine*, vol. 33, pp. 524–531, 4 2005.
- [72] R. H. Brophy, R. W. Wright, and M. J. Matava, “Cost Analysis of Converting from Single-Bundle to Double-Bundle Anterior Cruciate Ligament Reconstruction,” *The American Journal of Sports Medicine*, vol. 37, pp. 683–687, 2 2009.
- [73] N. A. Mall, P. N. Chalmers, M. Moric, M. J. Tanaka, B. J. Cole, B. R. Bach, and G. A. Paletta, “Incidence and Trends of Anterior Cruciate Ligament Reconstruction in the United States,” *The American Journal of Sports Medicine*, vol. 42, pp. 2363–2370, 8 2014.
- [74] H. B. Sigurðsson, Sveinsson, and K. Briem, “Timing, not magnitude, of force may explain sex-dependent risk of ACL injury,” *Knee surgery, sports traumatology, arthroscopy : official journal of the ESSKA*, vol. 26, pp. 2424–2429, 8 2018.
- [75] S. Beischer, E. Hamrin Senorski, C. Thomeé, K. Samuelsson, and R. Thomeé, “Young athletes return too early to knee-strenuous sport, without acceptable knee function after anterior cruciate ligament reconstruction,” *Knee surgery, sports traumatology, arthroscopy : official journal of the ESSKA*, vol. 26, pp. 1966–1974, 7 2018.
- [76] A. Kotsifaki, S. Van Rossom, R. Whiteley, V. Korakakis, R. Bahr, V. Sideris, and I. Jonkers, “Single leg vertical jump performance identifies knee function deficits at return to sport after ACL reconstruction in male athletes,” *British Journal of Sports Medicine*, pp. 2021–104692, 2 2022.
- [77] L. C. Schmitt, M. V. Paterno, and T. E. Hewett, “The Impact of Quadriceps Femoris Strength Asymmetry on Functional Performance at Return to Sport Following Anterior Cruciate Ligament Reconstruction,” *Journal of Orthopaedic & Sports Physical Therapy*, vol. 42, pp. 750–759, 9 2012.

- 
- [78] A. Phinyomark, C. Limsakul, and P. Phukpattaranont, "A Novel Feature Extraction for Robust EMG Pattern Recognition," *Journal of Computing*, vol. 1, pp. 71–80, 12 2009.
- [79] C. Spiewak, M. R. Islam, M. Assad-Uz-Zaman, and M. Rahman, "A Comprehensive Study on EMG Feature Extraction and Classifiers," *Open Access Journal of Biomedical Engineering and its Applications*, vol. 1, 2 2018.
- [80] A. Parnandi, J. Uddin, D. M. Nilsen, and H. M. Schambra, "The Pragmatic Classification of Upper Extremity Motion in Neurological Patients: A Primer," *Frontiers in neurology*, vol. 10, p. 996, 9 2019.
- [81] A. Ayman, O. Attalah, and H. Shaban, "An Efficient Human Activity Recognition Framework Based on Wearable IMU Wrist Sensors," in *2019 IEEE International Conference on Imaging Systems and Techniques (IST)*, pp. 1–5, 2019.
- [82] V. K. Ayyadevara, "Gradient boosting machine," in *Pro machine learning algorithms*, pp. 117–134, Springer, 2018.
- [83] C. Crema, A. Depari, A. Flammioni, E. Sisinni, T. Haslwanter, and S. Salzmann, "IMU-based solution for automatic detection and classification of exercises in the fitness scenario," in *2017 IEEE Sensors Applications Symposium (SAS)*, pp. 1–6, 2017.
- [84] P. Paul and T. George, "An effective approach for human activity recognition on smartphone," in *2015 IEEE International Conference on Engineering and Technology (ICETECH)*, pp. 1–3, 2015.
- [85] D. Meldrum, E. Cahalane, R. Conroy, D. Fitzgerald, and O. Hardiman, "Maximum voluntary isometric contraction: Reference values and clinical application," *Amyotrophic Lateral Sclerosis*, vol. 8, pp. 47–55, 1 2007.

- 
- [86] Y. Osawa, S. A. Studenski, and L. Ferrucci, “Knee extension rate of torque development and peak torque: associations with lower extremity function,” *Journal of Cachexia, Sarcopenia and Muscle*, vol. 9, pp. 530–539, 6 2018.
- [87] “Duchenne muscular dystrophy,” Mar 2021.
- [88] E. W. Ottesen, “Iss-n1 makes the first fda-approved drug for spinal muscular atrophy,” *Translational neuroscience*, vol. 8, no. 1, pp. 1–6, 2017.
- [89] S. D. Din, S. Patel, C. Cobelli, and P. Bonato, “Estimating fugl-meyer clinical scores in stroke survivors using wearable sensors,” *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 5839–5842, 2011.
- [90] H. Wang, Z. Wu, and E. P. Xing, “Removing confounding factors associated weights in deep neural networks improves the prediction accuracy for healthcare applications.,” in *PSB*, pp. 54–65, World Scientific, 2019.
- [91] M. Ziaeefard and R. Bergevin, “Semantic human activity recognition: A literature review,” *Pattern Recognition*, vol. 48, no. 8, pp. 2329–2345, 2015.
- [92] M. L. Pearl, F. van de Bunt, M. Pearl, N. Lightdale-Miric, S. Rethlefsen, and J. Loiselle, “Assessing shoulder motion in children: age limitations to mallet and abc loops,” *Clinical Orthopaedics and Related Research*®), vol. 472, no. 2, pp. 740–748, 2014.
- [93] M. G. Contesse, K. Davis, L. D. Pазze, C. McSherry, and M. G. Leffler, “Home-based versus in-clinic functional assessments: Preferences of caregivers and patients with duchenne.,” in *Parent Project Muscular Dystrophy*, 2020.
- [94] P. E. McKnight and J. Najab, “Mann-Whitney U Test,” in *The Corsini Encyclopedia of Psychology*, p. 1, John Wiley & Sons, Ltd, 2010.
- [95] V. Gritsenko, E. Dailey, N. Kyle, M. Taylor, S. Whittacre, and A. K. Swisher, “Feasibility of Using Low-Cost Motion Capture for Automated Screening of Shoulder Motion

- 
- Limitation after Breast Cancer Surgery,” *PloS one*, vol. 10, pp. e0128809–e0128809, 6 2015.
- [96] X. Chen, J. Siebourg-Polster, D. Wolf, C. Czech, U. Bonati, D. Fischer, O. Khwaja, and M. Strahm, “Feasibility of Using Microsoft Kinect to Assess Upper Limb Movement in Type III Spinal Muscular Atrophy Patients,” *PloS one*, vol. 12, pp. e0170472–e0170472, 1 2017.
- [97] A. Scano, F. Molteni, and L. Molinari Tosatti, “Low-Cost tracking systems allow fine biomechanical evaluation of upper-limb daily-life gestures in healthy people and post-stroke patients,” *Sensors*, vol. 19, no. 5, p. 1224, 2019.
- [98] I. Carpinella, D. Cattaneo, and M. Ferrarin, “Quantitative assessment of upper limb motor function in Multiple Sclerosis using an instrumented Action Research Arm Test,” *Journal of neuroengineering and rehabilitation*, vol. 11, no. 1, p. 67, 2014.
- [99] C. Wang, L. Peng, Z. Hou, J. Li, T. Zhang, and J. Zhao, “Quantitative Assessment of Upper-Limb Motor Function for Post-Stroke Rehabilitation Based on Motor Synergy Analysis and Multi-Modality Fusion,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 4, pp. 943–952, 2020.
- [100] J. Kim, H. Kim, and J. Kim, “Quantitative assessment test for upper-limb motor function by using EMG and kinematic analysis in the practice of occupational therapy,” in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 1158–1161, 2017.
- [101] J. M. Thuman, H. McMahon, P. Chow, M. Gerber, K. Dassoulas, L. Barnes, and C. A. Campbell, “The Performance of Patient-Worn Actigraphy Devices to Measure Recovery after Breast Reconstruction,” *Plastic and reconstructive surgery. Global open*, vol. 7, pp. e2503–e2503, 10 2019.

- 
- [102] Y. Liao, A. Vakanski, and M. Xian, “A Deep Learning Framework for Assessing Physical Rehabilitation Exercises,” *IEEE transactions on neural systems and rehabilitation engineering : a publication of the IEEE Engineering in Medicine and Biology Society*, vol. 28, pp. 468–477, 2 2020.
- [103] I. Mocanu, C. Marian, L. Rusu, and R. Arba, “A Kinect based adaptive exergame,” in *2016 IEEE 12th International Conference on Intelligent Computer Communication and Processing (ICCP)*, pp. 117–124, 2016.
- [104] S. McGraw, Y. Qian, J. Henne, J. Jarecki, K. Hobby, and W.-S. Yeh, “A qualitative study of perceptions of meaningful change in spinal muscular atrophy,” *BMC Neurology*, vol. 17, no. 1, p. 68, 2017.
- [105] R. L. Jones and M. Wallace, “Another bad day at the training ground: Coping with ambiguity in the coaching context,” *Sport, education and society*, vol. 10, no. 1, pp. 119–134, 2005.
- [106] T. Warren Liao, “Clustering of time series data—a survey,” *Pattern Recognition*, vol. 38, no. 11, pp. 1857–1874, 2005.
- [107] E. Scott, M. Eagle, A. Mayhew, J. Freeman, M. Main, J. Sheehan, A. Manzur, F. Muntoni, and T. N. S. C. N. f. P. N. Disease, “Development of a Functional Assessment Scale for Ambulatory Boys with Duchenne Muscular Dystrophy,” *Physiotherapy Research International*, vol. 17, pp. 101–109, 6 2012.
- [108] Y. Gloumakov, A. J. Spiers, and A. M. Dollar, “Dimensionality Reduction and Motion Clustering During Activities of Daily Living: Three-, Four-, and Seven-Degree-of-Freedom Arm Movements,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 12, pp. 2826–2836, 2020.

- 
- [109] M. Cortesi, A. Giovanardi, G. Gatta, A. L. Mangia, S. Bartolomei, and S. Fantozzi, “Inertial Sensors in Swimming: Detection of Stroke Phases through 3D Wrist Trajectory,” *Journal of sports science & medicine*, vol. 18, pp. 438–447, 8 2019.
- [110] G. B. D. Q. Davoli, J. Cardoso, G. C. Silva, R. d. F. C. Moreira, and A. C. Mattiello-Sverzut, “Instruments to assess upper-limb function in children and adolescents with neuromuscular diseases: a systematic review,” *Developmental Medicine & Child Neurology*, vol. 63, pp. 1030–1037, 9 2021.
- [111] D. D. Cvjetkovic, S. Bijeljic, S. Palija, G. Talic, T. N. Radulovic, M. G. Kosanovic, and S. Manojlovic, “Isokinetic testing in evaluation rehabilitation outcome after ACL reconstruction,” *Medical archives*, vol. 69, no. 1, p. 21, 2015.
- [112] C.-J. Su, C.-Y. Chiang, and J.-Y. Huang, “Kinect-enabled home-based rehabilitation system using Dynamic Time Warping and fuzzy logic,” *Applied Soft Computing*, vol. 22, pp. 652–666, 2014.
- [113] R. Cotton, “Kinematic Tracking of Rehabilitation Patients With Markerless Pose Estimation Fused with Wearable Inertial Sensors,” in *2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020) (FG)*, (Los Alamitos, CA, USA), pp. 588–594, IEEE Computer Society, 5 2020.
- [114] M. Causse, “Meeting report: 10th International Conference on Applied Human Factors and Ergonomics, 2019.,” *Aviation Psychology and Applied Human Factors*, vol. 10, no. 1, p. 44, 2020.
- [115] H. Jeong, S. S. Kwak, S. Sohn, J. Y. Lee, Y. J. Lee, M. K. O’Brien, Y. Park, R. Avila, J.-T. Kim, J.-Y. Yoo, M. Irie, H. Jang, W. Ouyang, N. Shawen, Y. J. Kang, S. S. Kim, A. Tzavelis, K. Lee, R. A. Andersen, Y. Huang, A. Jayaraman, M. M. Davis, T. Shanley, L. S. Wakschlag, S. Krogh-Jespersen, S. Xu, S. W. Ryan, R. L. Lieber, and J. A. Rogers, “Miniaturized wireless, skin-integrated sensor networks for quantifying

- 
- full-body movement behaviors and vital signs in infants,” *Proceedings of the National Academy of Sciences*, vol. 118, p. e2104925118, 10 2021.
- [116] H. Jia, Y. Wu, J. Liu, L. Yao, and W. Hu, “Mobile Golf Swing Tracking Using Deep Learning with Data Fusion: Poster Abstract,” in *Proceedings of the 17th Conference on Embedded Networked Sensor Systems*, SenSys '19, (New York, NY, USA), pp. 422–423, Association for Computing Machinery, 2019.
- [117] H. B. Mann and D. R. Whitney, “On a test of whether one of two random variables is stochastically larger than the other,” *The annals of mathematical statistics*, pp. 50–60, 1947.
- [118] M. Capecchi, M. G. Ceravolo, F. Ferracuti, S. Iarlori, V. Kyrki, S. Longhi, L. Romeo, and F. Verdini, “Physical rehabilitation exercises assessment based on hidden semi-markov model by kinect v2,” in *2016 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, pp. 256–259, IEEE, 2016.
- [119] Y. Liu, Q. Song, C. Li, X. Guan, and L. Ji, “Quantitative Assessment of Motor Function for Patients with a Stroke by an End-Effector Upper Limb Rehabilitation Robot,” *BioMed Research International*, vol. 2020, p. 5425741, 2020.
- [120] Y. T. Wang and H. P. Ma, “Real-Time Continuous Gesture Recognition with Wireless Wearable IMU Sensors,” in *2018 IEEE 20th International Conference on e-Health Networking, Applications and Services (Healthcom)*, pp. 1–6, 2018.
- [121] M.-C. Huang, W. Xu, Y. Su, B. Lange, C.-Y. Chang, and M. Sarrafzadeh, “Smart-Glove for Upper Extremities Rehabilitative Gaming Assessment,” in *Proceedings of the 5th International Conference on PErvasive Technologies Related to Assistive Environments*, PETRA '12, (New York, NY, USA), Association for Computing Machinery, 2012.

- 
- [122] G. Kwakkel, E. E. H. van Wegen, J. H. Burridge, C. J. Winstein, L. E. H. van Dokkum, M. Alt Murphy, M. F. Levin, and J. W. Krakauer, “Standardized Measurement of Quality of Upper Limb Movement After Stroke: Consensus-Based Core Recommendations From the Second Stroke Recovery and Rehabilitation Roundtable,” *Neurorehabilitation and Neural Repair*, vol. 33, pp. 951–958, 10 2019.
- [123] A. Filippeschi, N. Schmitz, M. Miezal, G. Bleser, E. Ruffaldi, and D. Stricker, “Survey of Motion Tracking Methods Based on Inertial Sensors: A Focus on Upper Limb Human Motion,” *Sensors (Basel, Switzerland)*, vol. 17, p. 1257, 6 2017.
- [124] H. Choi, Q. Wang, M. Toledo, P. Turaga, M. Buman, and A. Srivastava, “Temporal Alignment Improves Feature Quality: An Experiment on Activity Recognition with Accelerometer Data,” in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 462–4628, 2018.
- [125] M. Main, H. Kairon, E. Mercuri, and F. Muntoni, “The Hammersmith Functional Motor Scale for Children with Spinal Muscular Atrophy: a Scale to Test Ability and Monitor Progress in Children with Limited Ambulation,” *European Journal of Paediatric Neurology*, vol. 7, no. 4, pp. 155–159, 2003.
- [126] N. Nachar, “The Mann-Whitney U: A test for assessing whether two independent samples come from the same distribution,” *Tutorials in quantitative Methods for Psychology*, vol. 4, no. 1, pp. 13–20, 2008.
- [127] M. Barandas, D. Folgado, L. Fernandes, S. Santos, M. Abreu, P. Bota, H. Liu, T. Schultz, and H. Gamboa, “TSFEL: Time Series Feature Extraction Library,” *SoftwareX*, vol. 11, p. 100456, 2020.
- [128] A. Molina-Molina, E. J. Ruiz-Malagón, F. Carrillo-Pérez, L. E. Roche-Seruendo, M. Damas, O. Banos, and F. García-Pinillos, “Validation of mDurance, A Wearable

- 
- Surface Electromyography System for Muscle Activity Assessment,” *Frontiers in Physiology*, vol. 11, 2020.
- [129] R. S. McGinnis, J. B. Redrado, R. H. Choquette, B. D. Beynnon, J. R. Slauterbeck, T. W. Tourville, and M. J. Toth, “Wearable sensors capture differences in muscle activity and gait patterns during daily activity in patients recovering from ACL reconstruction,” in *2018 IEEE 15th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, pp. 38–41, 2018.
- [130] M. P. Ithurnburn, A. R. Altenburger, S. Thomas, T. E. Hewett, M. V. Paterno, and L. C. Schmitt, “Young athletes after ACL reconstruction with quadriceps strength asymmetry at the time of return-to-sport demonstrate decreased knee function 1 year later,” *Knee Surgery, Sports Traumatology, Arthroscopy*, vol. 26, no. 2, pp. 426–433, 2018.
- [131] “Fugl-meyer assessment upper extremity (fma-ue): Assessment of sensorimotor function,” Jul 2017.
- [132] Monirb, “Therapy lab image,” Aug 2014.
- [133] V. Vicon100, “Motion capture and motionbuilder demo,” Aug 2009.