Quantifying and Reversing Compensatory Movements by Persons Post-Stroke in the Ambient Setting

A Dissertation Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Aaron Miller

December 2021

© by Aaron Miller, 2021
All Rights Reserved.

To my wife, Amy Albaugh Miller. You have wholeheartedly embraced who I am. Where I am weak, you are my strength. Though I started us on this journey, you eventually became my guide. Without you, I am lost.

Acknowledgments

I would like to thank my advisor, Dr. Eric Espinoza-Wade, for his guidance during my time at the University of Tennessee. The standard to which you hold yourself in every aspect of life is an inspiration, and I consider myself lucky to have had you both as a mentor and role model for all these years. I promise to always write in LaTeX and pluralize data. I would like to thank Dr. Dustin Crouch and the Upper Limb Assist Lab for taking me in when I was a stray. Thanks to you, I always felt I had a place in Knoxville. I would also like to thank my other dissertation committee members, Dr. Jeff Reinbolt, Dr. Xiaoping Zhao, and Dr. Amir Sadovnik for their valuable input to my dissertation. My advisors and committee members always pushed me to do the best research of which I was capable.

I would like to thank my friends in the Behavior, Robotics and Sensing Group who were there through all the long edits and conference preparations. I will never forget all you have done for me, and I wish you the best in your life's journeys. I also have to thank Dr. Susan Duff and our other collaborators at Chapman University and Teacher's College, Columbia University for their tireless efforts in collecting data, writing, and editing. I also need to thank Dr. Aravind Sundararajan for his work on development of the application we used in this research and for all the advice and support he has given me over the years.

Thank you to my parents who gave me the opportunity to attend college in the first place and always supported my decisions as I figured out my path. Thank you to my in-laws, my second mom and dad, who were always willing to dog-sit when I needed a break or buy gluten-free pizza just to have me around. Thanks to my brothers, Jeremy and Michael, for contextualizing the quote, "It makes a considerable difference to me, having someone with me on whom I can thoroughly rely." I would also like to thank my gaming group, particularly Matt and Sunny, for providing me with a weekly opportunity to goof off and de-stress.

Finally, I have to thank Dr. Robert Minneci. I rolled a natural 20 when I met you buddy, and I can't imagine having done any of this without you.

Abstract

Nearly 800,000 people in the United States suffer stroke annually. Following the onset of stroke, survivors will exhibit deficits, such as hemiplegia, which will limit their function and ability to perform activities of daily living (ADLs). In order to regain independence, many stroke survivors will employ maladaptive compensatory strategies to help with the completion of tasks. Compensation is generally defined as any performance of a task that is different than the way it may have been performed before the onset of a neurodegenerative disorder. While for some severely impaired individuals, compensation may be necessary, for most these maladaptive strategies ultimately lead to a decreased quality of life and lifespan. Compensatory behaviors are most often developed in the ambient setting when an individual is unmonitored by a health care professional. Thus, it is necessary to develop a system which can autonomously detect compensatory behavior and then administer feedback or an intervention based on the compensation measured.

To develop this system, we first performed a scoping review to determine the specific, quantitative definitions of feedback described in the literature. The results of this analysis showed that compensation is most often defined in the context of a segment and a task being performed. Given that specific definitions of compensation are task specific, we used minimally intrusive inertial measurement units (IMUs) to evaluate the significant differences between commonly performed tasks (unimanual, bimanual symmetric, and bimanual asymmetric) and between groups (control and post-stroke). We then tested the capability of various machine learning methods to differentiate between groups and task types given features derived from the time-series human motion data. Ultimately, we created an echo state neural network (ESNN) which could differentiate between healthy and unhealthy task performance with an acceptable level of accuracy. We then developed an application

which could use the ESNN in real time to detect compensatory behaviors. Using language specifically designed to enhance recovery, a control participant was given feedback on task performance in a four hour pilot trial of the system, and the results of that analysis was reported.

The goal of this research was to aid in the long term recovery and decrease the use of compensatory strategies for persons post-stroke.

Table of Contents

1 Introduction **1**

2 Background **4**

2.1 Stroke 4

2.2 Function Vs. Recovery 5

2.3 Compensation 5

2.3.1 Nonuse Compensation 5

2.3.2 Behavioral Compensation 7

2.3.3 Motor Compensation 8

2.3.4 Other Compensation 9

2.4 Measuring Compensation 9

2.4.1 Non-intrusive Sensing Techniques 9

2.5 Feedback 10

2.5.1 Knowledge of Performance Vs. Knowledge of Results 11

2.5.2 Positive Vs. Negative Reinforcement 12

2.5.3 Frequency 12

2.6 Practical Application 13

2.7 Research Design 14

3 How do Researchers Quantify Upper Extremity Motor Compensation? A Scoping Review **17**

3.1 Introduction 17

3.2 Methods 19

3.2.1 **Identifying the research questions (Stage 1)** . . . 20
3.2.2 **Identifying relevant studies (Stage 2)** . . . 20
3.2.3 **Study selection (Stage 3)** . . . 21
3.2.4 **Charting data (Stage 4)** . . . 23
3.2.5 **Collating and summarizing the results (Stage 5)** . . . 23
3.3 Results . . . 24
3.3.1 Overview of included publications . . . 24
3.3.2 Explicit definitions of UE motor compensation . . . 26
3.3.3 Quantitative measures of UE motor compensation . . . 28
3.4 Discussion . . . 28
3.4.1 Explicit definitions of UE motor compensation . . . 28
3.4.2 Identify specific, quantitative measures of UE motor compensation assigned to each explicit definition (if given), and their prevalence of use in published literature. . . . 37
3.4.3 Determine any existing taxonomies for such compensation. . . . 38
3.5 Implications for Our Approach . . . 38

4 Novel Method to Quantify Interlimb Coordination in Persons with Hemiparesis Post-Stroke **40**
4.1 Introduction . . . 40
4.2 Methods . . . 43
4.2.1 Participants . . . 43
4.2.2 Clinical Characteristics and Demographics . . . 44
4.2.3 Experimental Procedure . . . 44
4.2.4 Data Processing . . . 47
4.2.5 Interlimb Coordination Measures . . . 47
4.2.6 Interlimb Coordination Algorithm . . . 50
4.2.7 Data Analyses . . . 54
4.3 Results . . . 54
4.4 Discussion . . . 61

4.4.1 Limitations . . . 62
4.4.2 Implications . . . 63

5 Comparison of Machine Learning Approaches for Classifying Upper Extremity Tasks in Individuals Post-Stroke 64

5.1 Introduction . . . 64
5.2 Data Collection and Processing . . . 66
5.2.1 Participants . . . 66
5.2.2 Design . . . 66
5.2.3 Procedure . . . 67
5.3 Supervised Learning . . . 67
5.3.1 Methods . . . 68
5.3.2 Results . . . 71
5.4 Data Exploration . . . 80
5.4.1 Methods . . . 80
5.4.2 Results . . . 83
5.5 Discussion . . . 87

6 Classifying Unimanual and Bimanual Upper Extremity Tasks in Individuals Post-Stroke 89

6.1 Introduction . . . 89
6.2 Methods . . . 91
6.2.1 Design . . . 91
6.2.2 Procedure . . . 91
6.2.3 Feature Extraction . . . 92
6.2.4 Data Preprocessing and Shaping . . . 94
6.2.5 Model Structure . . . 94
6.3 Results . . . 95
6.4 Discussion . . . 98
6.5 Acknowledgement . . . 99

7 Wearable Sensors for Reducing Compensation in Persons Post-Stroke at Home **100**

7.1 Introduction . . . 100

7.2 Methods . . . 103

7.2.1 Population . . . 107

7.2.2 Design . . . 107

7.3 Results . . . 108

7.4 Discussion . . . 110

7.5 Conclusion . . . 115

8 Conclusion **116**

8.1 Summary . . . 116

8.2 Data Characterization . . . 119

8.3 Performance of Machine Learning Approaches . . . 120

8.4 Neural Networks Performance . . . 121

8.5 Future Work . . . 123

Bibliography **124**

Appendices **160**

A Data Extraction Form . . . 161

B Data Separation by Band . . . 166

B.1 Methods . . . 166

B.2 Results . . . 168

B.3 Discussion . . . 173

Vita **174**

List of Tables

3.1 Organization of publications according to increasing specificity of compensation definition 29

3.2 Compensatory movements are specifically qualified a number of different ways across the papers we analyzed. This table shows the taxonomies that have been used to qualify compensation for each segment. The table is meant to be read [**Prefix**] **Segment** [**Suffix**], where the percentage following the suffix (or prefix if there is no suffix) is the percentage of papers which used this taxonomy to define a specific compensatory behavior. In instances where no prefix or suffix was used in the paper, that section of the table was left blank. In some instances, multiple prefixes preceded a single suffix, and in these cases, the percentages listed by the prefixes indicate the percentage of papers which used the listed prefix along with the listed suffix. The discrepancy between percentages listed for prefixes and suffixes in these cases is accounted for by the number of papers which used the suffix with no prefix. 30

4.1 Participant Characteristics 45

5.1 Average correctly classified instances when using only participants with certain FM scores and excluding those scores which were outliers: top 10 refers to the 10 participants whose FM scores were above 30, bottom 7 refers to the 7 participants whose scores were below 30, and middle 17 refers to a combination of the two aforementioned groups. 84

5.2 Correctly classified instances post-stroke and control data when separated by handedness. 85

6.1 Data synthesis approach 93
6.2 MLP Hyper-parameters 96

7.1 Feedback Conceptualization 105
7.2 Task performance divided by video and compared to predicted results 112
7.3 Feedback delivered during trial 113

1 Precision, recall, and F1 scores for preprossessing of control participants' data. 169
2 Precision, recall, and F1 scores for preprossessing of post-stroke participants' data. 170
3 Loss, Accuracy, Validation Loss, and Validation Accuracy of multi-layer perceptron analysis of preprocessed control data. 172

List of Figures

2.1 Following the onset of stroke, there are many paths a patient can take to functional recovery. Unfortunately, some of the more recently developed paths face roadblocks along the way, such as the interruption of a patient's normal routine or high cost associated with the therapy. 15

3.1 Flow chart depicting the article selection process 22
3.2 Publication years of selected articles 25
3.3 Segments (left) and tasks (right) for which compensatory behaviors are defined 27

4.1 Subjects wore 5 sensors on the sternum, L/R wrist, and L/R upper arm (A) during performance of 6 different tasks performed twice (B), including reaching for a water bottle as shown here; and performance of the Adult Assisting Hand Assessment Stroke (AdAHA-Stroke), which involved unwrapping and wrapping a present (C). 46
4.2 Raw acceleration data extracted from the left and right wrists of a control (left side plots) and persons post–stroke (right side plots) performing the task *Fold towel*. 48
4.3 The interlimb coordination algorithm utilizes amplitude, time, and frequency data from all five body-worn sensors to create an array of numerical values for unimanual, bimanual asymmetric, and bimanual symmetric gestures. . . 51
4.4 A) Sequence of algorithm construction and B) two examples. 52
4.5 Three key sensor to sensor comparisons: S-W1 (green line), W1-W2 (blue line), and W1-U1 (red line). 55

4.6 Similarity metrics for representative participants (Post-stroke – 19, 46 years old, UE-FMA-30; and Control 5, 49 years old) for the 10 sensor-to-sensor comparisons by task type: Unimanual (U), Bimanual Asymmetric (A), and Bimanual Symmetric (S). The comparisons were: a) sternum to wrist 1 (S-W1), b) wrist 1 to wrist 2 (W1-W2); and c) wrist 1 to upper arm 1 (W1-U1). Dark red color blocks indicate similarity (correlation) was closer to 1.0; dark blue color blocks indicate similarity closer to 0. 57

4.7 Mean (SEM) similarity metrics by group for each task type (unimanual , bimanual symmetric, and bimanual asymmetric) and the three sensor-to-sensor comparisons: Wrist to sternum (S-W1), wrist 1 to wrist 2 (W1-W1), and wrist 1 to upper arm 1 (W1-U1). 58

4.8 Significant differences from the within group, post-hoc analyses between the three different sensor locations and three different task types for healthy controls and persons post-stroke based on sensor comparisons. Red dots for controls and blue dots for persons post-stroke represent significant differences at $p<0.01$. 59

4.9 Arm Use Ratio based on the integral from wrist sensor data (paretic/non-paretic arm) for acceleration, angular rate of change, and orientation associated with A) UE Fugl Meyer Score; and B) Ad-AHA Stroke logit score. . . . 60

5.1 Correctly Classified Instances versus Number of Features Selected for Control (A) and Post-Stroke (B) Populations. The results are shown for the nine classifiers used, as well as an average of all nine classifiers. 72

5.2 Correctly Classified Instances versus Number of Features Selected for Control (A) and Post-Stroke (B) Populations. The average of the nine classifiers is shown for one and two second windows, as well as the full task length. 74

5.3 Correctly classified instances when hand-picked features were selected (A) compared to selecting features from the full extraction process (B). The nine classifiers and their averages are shown for both Post-Stroke and Control groups for each feature selection process. 75

5.4 Correctly classified instances as percent overlap of the two second window increased for Control (A) and Post-Stroke (B) groups. 76
5.5 Correctly classified instances for persons Post-Stroke and Controls when the training data was randomly divided and when the data was separated manually, leaving two participants out to test the model and train the remaining participants. 78
5.6 The Fugl-Meyer, which can be scored from 0 (worst performance) to 66 (best performance), is plotted versus the participant labels in the order the data was collected. The blue band represents the participants which are not outliers, the green band represents the lower portion of those participants which are not outliers, and the orange band represents the upper portion. 82
5.7 Exploration of the raw data. Data is shown in a histogram which demonstrates the normalized Area Under the Curve (AUC) for a given segment and task. The left column shows control participants and the right post-stroke; the rows represent the sternum, non-dominant/more affected wrist, the dominant/less affected wrist, non-dominant/more affected upper, and the dominant/less affected upper respectively. The color of the bars demonstrate the task being performed: blue for unimanual, red for bimanual asymmetric, and yellow for bimanual symmetric. 86

6.1 Accuracy (i), validation accuracy (ii), and test accuracy (iii) of the MLP classifier when using: (a) raw data; (b) raw data and the derived AUC feature; (c) raw data, AUC, and one other derived feature; or (d) raw data, AUC, and two other derived features for both the Control (**row 1**) and Post-Stroke (**row 2**) groups 97

7.1 Window for connecting and configuring devices and starting streaming (a) and the homepage (b) with options to **REHAPP**) return to screen (a); **FEEDBACK**) request feedback as shown at the top of the screen; **MY DATA**) see the screen shown in **Figure 7.2(a)**; or **SETTING**) adjust settings. 109

7.2 Results as shown in the application **(a)** and reorganized for clarity in the format of a pie chart **(b)**. 111

1 Area Under the Curve of the Less Affected Limb plotted versus the Area Under the Curve of the More Affected Limb for Control (a) and Post-Stroke (b) Groups. 167

2 Results of preprocessing the data by separating into bands. A classifier is used to sort data into a band (a); if the initial classification was incorrect, the data is then sorted to the correct band (b); finally, each band has a specific classifier which classifies the tasks for that band (c). 171

Chapter 1

Introduction

Of the nearly 800,000 people who suffer from stroke annually, more than sixty-three percent continue to suffer from chronic disability [267]. Hemiparetic stroke leaves one side of a patient's body weakened, decreasing the patient's ability to perform activities of daily living and maintain functional independence [128]. This creates a need for more substantial rehabilitative techniques, in the clinical, lab, and home setting. Because of the limitations of current rehabilitation, many rehabilitative techniques rely on restoring a patient's function. Restoring function means that a patient will be able to complete a task given any tools at his or her disposal unlike full recovery which means that the patient will perform a task the way her or she would before the onset of stroke [50]

Patients post-stroke will often use compensatory techniques to regain function. Compensation is the performance of a task differently due to injury or disability [123]. While for some patients who are severely impaired, compensatory techniques may be necessary, often these compensatory mechanisms can be harmful in the long term, reducing quality of life and longevity. There are several types of compensatory mechanisms that patients post-stroke will use, but the most prevalent are non-use of the affected limb and trunk compensation during motor reaching tasks. The method for dealing with non-use is constraint induced movement therapy (CIMT), which is a method that forces use of the affected limb through functional task training and restriction of the less affected limb [127], and the method for dealing with trunk engagement during reaching is trunk restraint [175]. These methods help with patient recovery, but they also limit a patient's function.

Compensatory techniques are primarily believed to be developed in the ambient setting when a patient is not under the supervision of a clinician or therapist [177]. Because of this, research has been trending toward automated detection of compensatory techniques [199]. Using tethered, motion analysis in the laboratory setting, it is possible to define, measure and quantify compensation [292]. However, it is more difficult to measure compensation in the ambient setting where motion analysis is not as precise and the motions being performed are varied. A favored method of in-home analysis is using tri-axial accelerometers strapped to the limbs to monitor motion [57]. Using this technique has proven useful in measuring amount of use, but has difficulty distinguishing between recovered movement and compensatory behaviors.

Many existing rehabilitative techniques aim to reduce a patient's impairment, but do so at the cost of function. Delivering feedback to a participant regarding the quality of the motion as a task is performed requires the patient to constantly thinking about recovery, without necessarily limiting function. We have named this alternative to CIMT **F**eedback **I**nduced **M**ovement **T**herapy (**FIMT**).

There are several types of feedback that can be delivered to a patient. It is important that feedback delivered enhances a participants expectations of results, increases a participants autonomy, and externalizes a participants focus [290]. Feedback must also be delivered at a the right time and frequency. If feedback is delivered too little, then participants will see limited reduction of impairment, and if it is delivered too often, participants run the risk of becoming fatigued with the feedback.

Using these new techniques, now more than ever clinicians can develop personalized rehabilitation strategies for their patients [2]. However, stroke disproportionately affects disadvantaged groups [71], and personalized healthcare is largely restricted to the educated, healthy, and wealthy [156]. Although minimally intrusive sensing techniques are an improvement on the unobtainably expensive motion analysis equipment that is used in the laboratory setting, there is still a need for low cost home rehabilitation devices [187]. Some methods to help offset the cost of personalized at-home rehabilitation are to limit the number of devices used or to use a device that has function outside of rehabilitation.

In this study, we evaluate the scope of state-of-the-art literature on compensation to quantitatively define specific compensatory behaviors and develop a comprehensive taxonomy for compensatory behavior (**Chapter 3**). Next, we aim to use features extracted from tri-axial inertial measurement units to differentiate between group and task performance of activities of daily living (**Chapter 4**). Then we explore methods for classifying task performance (**Chapter 5**) and develop an echo state neural network (ESNN) that can differentiate between task types and groups in the ambient setting (**Chapter 6**). Then using that classifier, we developed an application that can deliver feedback to participants in real–time and test the system with a control participant in a proof of concept trial (**Chapter 7**).

Chapter 2

Background

2.1 Stroke

In the United States, nearly 800,000 people suffer from stroke each year, and of those, nearly 140,000 will result in the death of the stroke patient. One in twenty deaths in the United States is attributed to stroke, ranking it fifth among all causes of death, behind diseases of the heart, cancer, Chronic Lower-Respitory Disease (CLRD), and unintentional injuries/accidents [183]. However, since the year 2000, the number of stroke survivors has increased by 38 percent. This means that there are more people living with the long-term effects of stroke, making it the leading cause of serious long-term disability [28].

This study will focus on patients who have suffered hemiplegic or hemiparetic stroke. A patient who has suffered this type of stroke will experience a loss of strength in the arm, leg, and sometimes face on one side of the body. Hemiplegia refers to a severe or complete loss of strength, whereas hemiparesis refers to a mild loss of strength [128]. As a result of this loss in strength on one side of the body, researchers who examine the effects of stroke on the upper body will refer to the more affected limb as the paretic limb and the less affected limb as the nonparetic limb.The loss of function on the paretic side that results from hemiparetic stroke can significantly decrease a patient's quality of life and his/her ability to perform the simple activities of daily living (ADLs) that allow them to be independent [108, 46]. Inability to perform tasks with the affected limb often forces a patient to use compensatory strategies to maintain his/her independence and ability to perform simple tasks.

2.2 Function Vs. Recovery

Following a stroke, there is a difference between a patient's function (being able to perform tasks using all method's at one's disposal) and recovery (ability to perform tasks as one would before stroke) [50, 123]. The difference between these two is that recovery necessitates a reduction of a participants impairment, where functionality does not. Impairment may be described as a modification of body structures such as a significant deviation in structure of the nervous system or structures related to movement, for example the arm and/or hand. Common motor impairments following stroke include: paresis, loss of fractionated movement, abnormal muscle tone, and/or changes in somatosensation [129]. Although the goal following stroke would be full recovery, it can be complicated to achieve because impairment is not static, the nature of the impairment may change, and a patient post-stroke may exhibit multiple impairments simultaneously [205]. Given these complications, patients post-stroke may be encouraged to use compensatory techniques to complete tasks.

2.3 Compensation

Generally defined, compensation is using one or more motor strategies alternative to those used by an unimpaired individual. [123, 137] Given the severity of an individual's impairment, there are circumstances under which a patient is better off using compensatory techniques. For some individuals, the use of compensatory techniques is the only way to regain any functional independence. However, these techniques can be detrimental if they are used over long periods of time, leading to a decreased lifespan and long-term quality of life [35, 146, 6, 174]. For the purposes of this research, compensation can be divided into four categories: nonuse of the affected limb, behavioral compensation, motor compensation, and other compensation that does not fit into the other four categories.

2.3.1 Nonuse Compensation

Nonuse of the affected limb is one of the most common, and subsequently one of the most studied, forms of compensation. After the onset of hemiparetic stroke, the patient loses

function in the affected limb, and one of the most simple methods of compensation is nonuse of the affected limb. Because the brain is not static, when the paretic limb is not being used following stroke, the brain begins to rewrite the neural pathways that were dedicated to the control of the unused limb. Given that the other body segments are being required to perform additional and more complex tasks than before, more resources are wired to controlling the functional body segments. This leads to further impaired use of the affected limb, which leads to more disuse of the limb, which leads to more rewiring of the neural pathways in the brain. The result of this cycle is that the patient has not only trained him/herself not to use the limb but has also weakened his/her ability to do so even if they tried [241, 189, 238]. This cycle is known as learned nonuse. Studies have shown that forced use of the affected limb can assist in recovery. Often, to force a patient to use the affected limb, a rehabilitative approach known as constraint induced movement therapy is used.

Constraint Induced Movement Therapy

There have been several attempts to reverse learned nonuse, but by far the most prevalent therapy used to treat upper extremity learned nonuse is Constraint Induced Movement Therapy (CIMT). CIMT is a method by which the nonparetic or dominant limb is constrained forcing the patient to use the paretic or non-dominant limb. Typically, this is by having the participant wear a mitt that does not permit use of the less affected limb [127]. By constraining the limb in this way for approximately 90 % of waking hours, the participant is forced to use the more affected limb, thus reducing physical impairment and inducing plastic changes in the brain. CIMT also involves intensive practice of the paretic limb that is designed to enhance task specific use and methods to transfer the gains achieved in the clinical setting to the ambient setting. Finally, CIMT includes diary-keeping or goal setting, sometimes in the form of a motor activity log (MAL) [285, 284, 54], which may be more important than the actual motor practices. The combination of these practices has been proven to increase recovery rates, improve rehabilitation, and reduce the effects of learned nonuse [60, 143, 58]. The downside of this therapeutic technique is that it makes it difficult to perform normal tasks early in recovery and performing two-handed tasks becomes impossible. CIMT can also be time consuming to administer, and the inability to perform ADLs in the

acute stages after stroke can limit a patient's functional independence. It can also have the effect of instigating learned nonuse of the less affected or non-paretic limb [282].

The task then becomes to gain the benefits of CIMT without subjecting the patient to the subsequent negative effects. Several different forms of therapy have been devised as an improvement on CIMT. Modified Constraint Induced Movement Therapy (mCIMT) varies in dosage, timing, and composition of therapy. mCIMT has been shown to have a positive effect on motor function of the paretic arm, muscle tone, arm-hand activities, self-reported amount of arm-hand use in daily life, self-reported quality of arm-hand movement in daily life, and basic ADLs [127]. mCIMT protocol recommends that treatment sessions last from 30 minutes to 6 hours a day, and from 2 to 7 sessions a week, for 2 to 12 weeks. Although this training is less extensive than the CIMT protocol, there remains a need for a training protocol that is less intrusive into the life of a patient post-stroke.

2.3.2 Behavioral Compensation

The second from of compensation that we have identified is behavioral compensation. Behavioral compensation occurs when someone engages in behaviors that allow them to avoid using the affected limb in a way that they would perceive as being difficult or impossible, including recruiting help to perform a task. Following stroke, it is not uncommon for a stroke survivor to avoid in engaging in meaningful activity or specific ADLs to avoid using the more affected limb [4, 19, 133]. An example of this would be a patient post-stroke asking someone cut up food at a restaurant because it is difficult to do without use of the affected limb, or even avoiding going to a restaurant because it would be difficult to avoid having to perform complicated or difficult tasks.

Although using this compensatory technique limits long-term recovery, it can be difficult to monitor and treat. Often, monitoring behavioral compensation comes in the form of self-reporting, which can be subject to bias [4, 19, 133]. Treatment of behavioral compensation is also difficult because behavioral compensation is often a symptom of underlying impairments.

2.3.3 Motor Compensation

The third form of compensation is motor compensation. Typically, motor compensation is defined as an inappropriate recruitment of body segments to complete a task, though this is not always the case. One motor compensatory strategy used by stroke patients is the fixation of certain body segments which limits the degrees of freedom available to complete a task. Another common form of motor compensation is trunk involvement during a reaching task [136], or a task where a subject is asked to grasp or point to an object outside of the neutral zone requiring elbow extension and shoulder flexion. During reaching tasks, trunk compensation has been significantly correlated with the level of motor impairment in participants post-stroke [50]. Increased trunk involvement coincides with a patient's impaired ability to extend the elbow, flex the shoulder, horizontally adduct the should, or perform any combination of these three movements. Other forms of motor compensation include inappropriate trunk inclination, flexion, rotation, or movement; inappropriate shoulder flexion, extension, abduction, or movement; inappropriate elbow flexion, extension, or movement; inappropriate flexion of the finger; hand posture; movement times; movement trajectories; movement velocities; reach distance; and changes in interlimb coordination [14, 15, 19, 27, 34, 50, 55, 102, 145, 146, 157, 158, 168, 169, 173, 175, 176, 177, 206, 208, 212, 213, 244, 245, 246, 263, 281, 293].

Interlimb Coordination

One important feature of upper extremity (UE) motor compensation is interlimb coordination (ILC) [265]. Interlimb coordination is critical to the performance of ADLs, and is known to be reduced in persons with post–stroke hemiparesis. Further, the ability to perform coordinated UE tasks is an important rehabilitation outcome that is often neglected according to patients and clinicians [20, 21]. Decreased upper extremity inter and intralimb coordination is indicative of maladaptive compensatory movement strategies, including inappropriate recruitment of limbs or segments to complete tasks. By evaluating the quantifiable differences between these task types, it is possible to use interlimb coordination as a proxy for recovery [113] (or a the reduction of compensatory behaviors).

2.3.4 Other Compensation

The final form of compensation considered in this study is other compensatory techniques that a patient uses to complete a task. Examples include the use of an exoskeletal brace that can compensate for the weight of the arm caused by gravity [96, 122], or an exoskeletal robotic thumb that can be used as a tool for grasping [99, 100]. Often, these other forms of compensation are not inherent to the participant and are used under the supervision of a researcher or clinician. Although such tools can be useful in increasing function following a stroke, they do not address the underlying cause of that decreased function, which is the impairment of the affected side of the body.

2.4 Measuring Compensation

Reinforcement of compensatory behaviors is thought to primarily occur in ambient settings when an individual is unmonitored by a health care specialist [177]. This has motivated the development of various sensor–based approaches to monitor motion in individuals with motor dysfunction [257, 243, 199]. Existing literature on compensation and ILC has largely focused on the evaluation of gait [57, 231, 247], or the use of tethered marker–based techniques unsuitable to ambient settings [216, 292]. Unfortunately, these existing research approaches that evaluate compensation or ILC using kinematic metrics are inappropriate for ambient monitoring given the requirement that users are stationary or tethered to measurement equipment.

2.4.1 Non-intrusive Sensing Techniques

It can be difficult to analyze upper extremity tasks outside of the laboratory environment. Because of the varied and under-analyzed nature of these tasks, many machine learning and pattern recognition approaches may not be suited for quantification and classification of ADLs. Manual tasks are not regularly repeated, and they involve multiple degrees of freedom meaning they can be accomplished using a multitude of strategies [6]. Because of this, in-home monitoring of patients post-stroke has remained a challenge.

An example of non-intrusive sensing techniques that have been proven useful for measuring amount (or quantity) of motion in the home are tri-axial accelerometers [257, 272]. By having patient's wear accelerometers, researchers have been able to reliably monitor and measure paretic limb use and to index upper extremity function in an ambient setting [243, 254, 259]. The use of these sensor based techniques provide daily information to the patient and are helpful in providing the clinician with insight into the patient's daily life as they plan interventions [148]. Sensor based **feedback** from accelerometers has been shown to increase self-efficacy and physical activity in patients post-stroke in the clinical or research setting, albeit in the lower extremity [77, 111, 121, 161]. Use of these sensors has also been shown to increase a clinician's knowledge of a patient's mobility related activity [64].

2.5 Feedback

Though measuring compensatory behaviors in ambient settings is critical, the ability to sense alone is insufficient. A method that has proven successful in motivating improved task performance is augmented feedback. In motor learning and control, the term augmented feedback is used to describe feedback that is received outside of the inherent feedback that is accessed when performing a task through visual, tactile, proprioceptive, and other sensory modalities [218]. Augmented feedback is used to build or strengthen and internal task model and to better facilitate learning. Providing feedback to a patient using minimally intrusive techniques can aid in nudging a patient toward recovery without limiting function.

This **F**eedback **I**nduced **M**ovement **T**herapy (**FIMT**) operates by using sensors to detect when compensatory behaviors, such as nonuse of the affected limb, occur. When compensation occurs, feedback is automatically sent to the patient to incentivize use of the affected limb rather than forcing use, which can be a method for revealing a patient's inherent movement capability even if it is less preferred by the patient following stroke [17, 148, 166]. The difference between an intervention such as FIMT when compared to CIMT for patients post-stroke, including those with high reliance on the less affected limb, is that, where CIMT can impede function by limiting movement, FIMT only promotes use of the more affected

limb without limiting function. When providing feedback to patients in this way, there are two major considerations: what type of feedback is being given and how often is the feedback being given?

2.5.1 Knowledge of Performance Vs. Knowledge of Results

There are multiple variables associated with the delivery of augmented feedback. One dimension includes the aspect of task performance, and can be separated into 'knowledge of performance' and 'knowledge of results.' The first of the two categories for feedback can either be knowledge of performance (KP) or knowledge of results (KR). Knowledge of performance is feedback delivered in a way that a participant understands how he or she performed a task. KP is typically verbal and delivered terminally, dealing with the aspects of task performance that are made evident as a task is being performed, such as the body and details of the movement. An example of this would sound like, "You successfully flexed your shoulder and extended your elbow." KP is often considered the humanized form of feedback and it is often the type of feedback that a clinician or therapist will deliver during the rehabilitation period. This feedback has been more effective in reducing impairment than a KR-based intervention in a laboratory setting [51].

Knowledge of results is also typically verbal and delivered terminally, but instead focuses on the task outcome. An example of this is, "You successfully moved your hand to the target." This feedback is more mechanical, more automated, and as such, it is much easier to for a machine to deliver this type of feedback during rehabilitation exercises. Because of this KR has been extensively studied in rehabilitation research [51, 279]. The ease of provision of KR grants itself to the type of automated, in-home information that can be quantified by sensors.

Both KP and KR incentivize performance (e.g., participants desire to hear good feedback regarding task performance) and provide motivation (e.g., the provision of feedback keeps participants focused on the tasks that will lead to rehabilitation). Participants may relate better to the contents of knowledge of performance feedback as it may be more comprehensible. It seems to be unclear which type of augmented feedback will be the most advantageous in the home setting.

2.5.2 Positive Vs. Negative Reinforcement

In addition to evaluating the effects that delivering feedback in the form of knowledge of performance versus knowledge of results has on recovery, recent research has increased focus on positive versus negative feedback. Positive, or rewarding, feedback is delivered when a patient or participant does something right as encouragement to continue the correct behavior while negative feedback is delivered when a person does something wrong to discourage incorrect behavior. Participants will respond better when given positive feedback, and three best practices as described by Lewthwaite et al. for delivering positive feedback are: enhancing expectancy of results, providing autonomy, and externalizing focus [290, 62, 49, 289].

Enhanced **E**xpectancies (**EE**) is a motivational variable that has been shown to promote learning. Examples of EE include: a participant receiving feedback only after performing well during a trial [13, 44, 52, 217]; being told, either truthfully or falsely, that his or her performance is better than the performance of peers [11, 139]; being told that peers typically perform well on the given task [288]; or increasing the participant's perception of success during training [43, 288, 196, 251]. **A**utonomy (**AS**) is another motivational variable that is important for optimal learning. Allowing participants to control various aspects of practice in motor learning (delivery of feedback [104], use of assistive devices [84], extent of practice [203], and frequency of skill demonstration [291]) has been shown to result in more effective learning. The final variable, **E**xternal **F**ocus (**EF**), is attentional and directs the participant's focus to the intended movement effect. This has been shown to enhance learning when compared to feedback that has an internal focus on body movements or not instructed focus [150, 164, 287]. Use of these positive reinforcement strategies have been shown to be key in optimizing motor recovery following stroke.

2.5.3 Frequency

With the ease of access that patients now have to information and communication, it can be easy to get quick and concise feedback on any number of questions, thoughts, or actions [126]. However, there is evidence that suggests that there is a time, or periods of time, where a

patient is more likely to receive feedback and respond positively [185]. It is also suggested that a patient who receives feedback too often, or with too high a frequency, will not respond as positively to the accumulating flow of information, developing **intervention fatigue** with the feedback, or a weariness associated with the intervention [87].

The time in which a patient gets feedback following the performance of a task is key to success of recovery. Recent work, examining **J**ust-**I**n-**T**ime **A**daptive **I**ntervention (**JITAI**), which is an intervention design that aims "to provide the right type (or amount) of support, at the right time, while eliminating support provision that is not beneficial," suggests that intervention adherence and retention are dependent on two important factors: engagement and fatigue [185]. Intervention engagement is defined as a "state of motivational commitment or investment in the client role over the treatment process" [117], and intervention fatigue is defined as a "state of emotional or cognitive weariness associated with intervention engagement" [185, 87].

If it can be assumed that patients participating the trial will demonstrate intervention engagement, then testing needs to be done to measure and mitigate intervention fatigue by adjusting the frequency of the feedback that is provided. If it were possible to autonomously detect every instance of compensatory behavior in the ambient setting, it is important to determine the number of instances of compensation for which feedback would be provided. As such, current research is employing these strategies to deliver the appropriate amount of feedback to the patient at the appropriate times [185, 248].

2.6 Practical Application

The ability to observe in home behaviors with sensors now makes it possible to generate personalized rehabilitation strategies tailored to each individual patient [2]. Research suggests that emerging technologies can facilitate reduction in disease risk and changes in behavior and that the impact of these interventions is increased when coupled with other engagement strategies [94, 204, 32]. However, while noncommunicable diseases disproportionately affect disadvantaged groups [71, 190, 156], the adoption of personalized health technologies is largely restricted to the educated, healthy and wealthy [200, 72, 66].

For lower-income populations, these devices remain largely unaffordable and inaccessible, and many patients with lower-income, or patients post-stroke, do not have, or maintain, the technological experience needed to adjust their behaviors based on the information being conveyed by these devices [67, 211, 187].

These sensor based methods for monitoring and measuring patient activity decrease the impact on patient life when compared to CIMT and have managed to decrease nonuse and perceived nonuse of the affected limb in patients post-stroke [277]. However, they do not address the cost associated with obtaining multiple sensors that are solely used for data collection. It is to the patient's benefit that the device(s) being used to facilitate FIMT is/are both user-friendly and low-cost, or at least provides enough function outside of data collection and rehabilitation to justify the cost, and one way to achieve this is to use commercially available devices. These commercially available devices are known as **C**ommercial-**O**ff-**T**he-**S**helf (**COTS**) wearables, and can be used in a variety of applications [110]. For a population post-stroke, an ideal intervention system is one with only one COTS wearable sensor collecting sufficient data on the affected limb such that it can detect nonuse or compensatory motor strategies and provide feedback to a patient in real-time.

2.7 Research Design

In this paper, we look at ways to categorize task performance, recognize compensatory behavior, and provide feedback to participants, in real-time, using minimally intrusive sensing techniques. We then provide feedback in a timely manner such that actionable change is both desirable and achievable. To achieve this, we first explored the ways that compensation is quantified in the form of a scoping review (**Chapter 3**), and we discovered that compensation is quantified in the context of a segment being used and a task being performed. We then explored using minimally intrusive sensing techniques to differentiate between tasks being performed and groups (Control and Post-stroke) performing the tasks (**Chapter 4**). We then evaluated the use of several machine learning methods to autonomously classify tasks and group performance (**Chapter 5**) before settling on a neural network which was capable of classifying chaotic, human-derived, time-series data

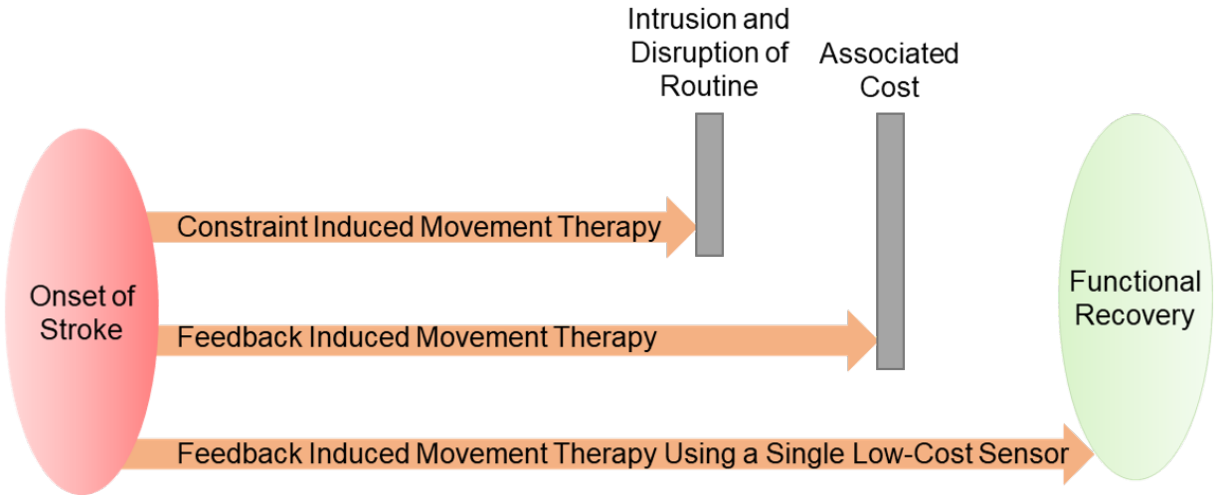


Figure 2.1: Following the onset of stroke, there are many paths a patient can take to functional recovery. Unfortunately, some of the more recently developed paths face roadblocks along the way, such as the interruption of a patient's normal routine or high cost associated with the therapy.

(**Chapter 6**). Finally, we developed an android application which used the neural network to autonomously distinguish between healthy and compensatory performance of tasks in persons post-stroke and controls in the ambient setting and provided feedback to participants to promote healthy task performance (**Chapter 7**). All of this work was done to reduce the use of compensatory behaviors in persons post-stroke in order to increase quality of life and lifespan.

Chapter 3

How do Researchers Quantify Upper Extremity Motor Compensation? A Scoping Review

3.1 Introduction

Nearly 800,000 people in the United States experience a stroke annually [267]. Many stroke survivors experience deficits in motor abilities that can lead to decreased independence [128], quality of life, and lifespan [82, 130, 280]. Often, stroke survivors will use compensatory movement strategies such as non-use of the affected limb, atypical movement, or coupling of the limb(s) to compensate for weakness, rigidity, and limited motor control [137, 265, 174] in order to regain function and independence. For severely impaired individuals, clinicians may support the use of compensatory techniques as it can be the only way to regain any functional independence; however, these compensatory strategies are maladaptive movements that can lead to long-term negative effects [241, 177, 35, 146, 6, 174]. Such compensatory behaviors are thought to evolve in ambient settings (as opposed to in clinical settings where individuals are monitored by care professionals) [177]. Because of this, recent work has been done to autonomously evaluate, record, and reverse compensation in the ambient setting [177]. Our goal is to determine the extent to which existing devices have been used

to autonomously quantify gestures and compensatory movements so that these maladaptive movement strategies can be addressed and reversed.

Clinical measures such as the Fugl-Meyer (FM) or the Wolf Motor Function Test (WMFT) are used to evaluate impairment and function. These measures score people down when compensatory strategies are used to complete a task; therefore, these tools have been used as proxies for quantifying compensation. However, these clinical tools are not designed to distinguish between function and recovery [92, 61, 74, 73] and are inappropriate as measures of motor compensation. Kinematic measures have been developed for differentiating between function and recovery, but these are primarily focused on specific, in-lab reaching tasks [232, 214, 270]. It is not realistic to use 3D kinematics for all patients or in home due to cost, time and personnel restrictions [22].

One method currently used to measure compensatory behavior in the ambient setting is the motor activity log (MAL) [261, 42]. The MAL is a questionnaire which asks participants to rate arm use during specific tasks on a discrete scale based on pre-stroke task performance. While the MAL has been shown to reliably measure real-world upper-extremity quality of motion in persons post-stroke, the tool allows for participants to complete tasks using compensatory behaviors. Thus, MAL scores are not always reflective of motor recovery. Finally, the nature of the assessment dictates that it cannot be used in real-time and that it can be influenced by patient objectivity, comprehension, and recall.

To evaluate behavior in real-time, minimally intrusive sensing techniques, such as accelerometry [130], have been used to autonomously differentiate functional and non-functional movement in the ambient setting [258]. These sensing techniques have also been used to evaluate arm-use following the onset of stroke [131, 86]. Currently, measuring functional movement or arm use is the primary method for autonomously describing real-time compensatory behaviors.

However, existing descriptions of such behaviors may be insufficient for determining specific limbs, segments, and joints involved in this behavior. Our pursuit of quantitative tools for measuring motor compensation revealed the need for examples of motor behaviors useful for the training and development of computational models. This revealed a dearth of specific, quantifiable, limb-based metrics of motor compensation. The purpose of the

scoping review is to address the extent and robustness of evidence on quantitative descriptors of compensatory motor behavior. Our research question is focused on the extent to which upper-extremity motor compensation has been quantified in individuals post-stroke, and what taxonomies (if any) exist with respect to this population. Therefore, the rationale for the scoping review is to explore the robustness of quantitative descriptions of upper extremity motor compensation.

Our goal is to determine existing quantifiable metrics for measuring, detecting, tracking, or quantifying upper-extremity motor compensation. This exploration will provide insight into the long-term quantification of ambient recovery responsiveness of motor behaviors to injury and intervention of in individuals post-stroke. Our objectives are therefore to:

1. Investigate the quantitative description of UE compensation in clinical literature.

2. Identify specific, quantitative measures of UE motor compensation, and their prevalence of use in published literature.

3. Determine any existing taxonomies for such compensation.

3.2 Methods

Our investigation of the literature was performed using a scoping review. A scoping review is a type of knowledge synthesis that focuses on an exploratory research question that queries an area or field of research of key concepts by searching the literature [53]. Scoping reviews are performed to examine the extent, range, and nature of research activity to summarize and disseminate research findings and to identify gaps in the existing literature [7, 135]. Our review will follow the updated recommendations to the original methodological framework [7] proposed by Levac et al. in [135]. This methodological framework consists of six stages: 1) Identifying the research question; 2) Identifying relevant studies; 3) Study selection; 4) Charting the data; 5) Collating, summarizing, and reporting results, and 6) Consultation (optional). The scoping review also follows the Preferred Reporting Items for Systematic reviews and Meta-Analysis extension for Scoping Reviews

(PRISMA-ScR) reporting guidelines [252]. The protocol was registered under the Open Science Framework (OSF).

3.2.1 Identifying the research questions (Stage 1)

Our research questions include:

1. **Research Question 1:** How have studies of UE impairment and function in clinical literature specifically described/defined compensation and the use of compensation following stroke?

2. **Research Question 2:** What specific quantitative measures of motor compensation are used in published literature?

3. **Research Question 3:** What taxonomies are assigned to the specific quantitative measures of compensation defined in published literature?

3.2.2 Identifying relevant studies (Stage 2)

Information sources: The databases used to conduct the search were PubMed, Cochrane, CINAHL Complete, Web of Science, PsycINFO, Scopus, and IEEE Xplore. We did not have any contact with authors to identify additional sources. The final database searches all took place on the same day (June 3^{rd}, 2019).

Selected keywords included in the search were: (stroke OR post-stroke OR TIA OR Transient Ischemic Attack OR CVA OR Cerebrovascular accident OR capsular infarction OR hemorrhagic stroke OR ischemic OR ischemic stroke) AND (compensation* OR compensatory behavior OR compensatory technique* OR compensatory strateg* OR compensatory motor strateg* OR compensatory response OR compensatory mechanisms OR alternate movement OR substitution OR adaptation OR redundancy) AND (rehabilitation OR recovery OR rehabilitative therapy OR chronic OR long-term OR restoration OR repair OR restitution) AND (upper limb OR upper arm OR upper extremity OR upper body OR upper arm OR trunk OR back).

All search results were exported to EndNote (EndNote X9, Philadelphia, PA). Publications were then excluded if they did not meet the aforementioned criteria regarding language or type of publication. Publications were then further evaluated to determine whether they met three inclusion criteria: Does the article/study examine stroke, the effects of stroke, or stroke rehabilitation? Does the article examine upper extremity? Does the article/study refer to compensation/compensatory behavior? If the three inclusion criteria were met, then the publication was evaluated using a data extraction form.

3.2.3 Study selection (Stage 3)

Eligibility criteria: Articles were included in this scoping review if they met the following inclusion criteria: 1) The study population consisted of individuals who have experienced a cerebrovascular accident (CVA)/stroke; 2) The focus of the study was on the upper extremity; and 3) The study described the problem of, an intervention to address, or the avoidance of upper extremity compensatory motor behavior; 4) Publications written in English only 5) Chapters, books, theses, conference proceedings $>$ 1 page in length.

Selection of sources of evidence: Returned search results were compiled in EndNote and duplicates were removed. EndNote search function was used to download copies of each article for review. Each article was screened to determine if it met the three inclusion criteria.

Figure 3.1 outlines the study selection process. The search resulted in over 1500 articles from the 8 databases. Thirteen-hundred articles then remained after removing redundancies. An initial screening of abstracts was performed by to remove any obviously inappropriate articles. This included articles which were not analyses of stroke, articles which did not evaluate the upper extremity, and articles which did not evaluate compensatory movement strategies. Abstracts, one-page conference proceedings, articles not in English, etc. were also removed. The remaining 800 abstracts more closely screened, which resulted in the removal of an additional 400 documents. The final 400 were carefully evaluated to determine whether they specifically described compensation post-stroke, resulting in the final 94 articles that fit the inclusion criteria.

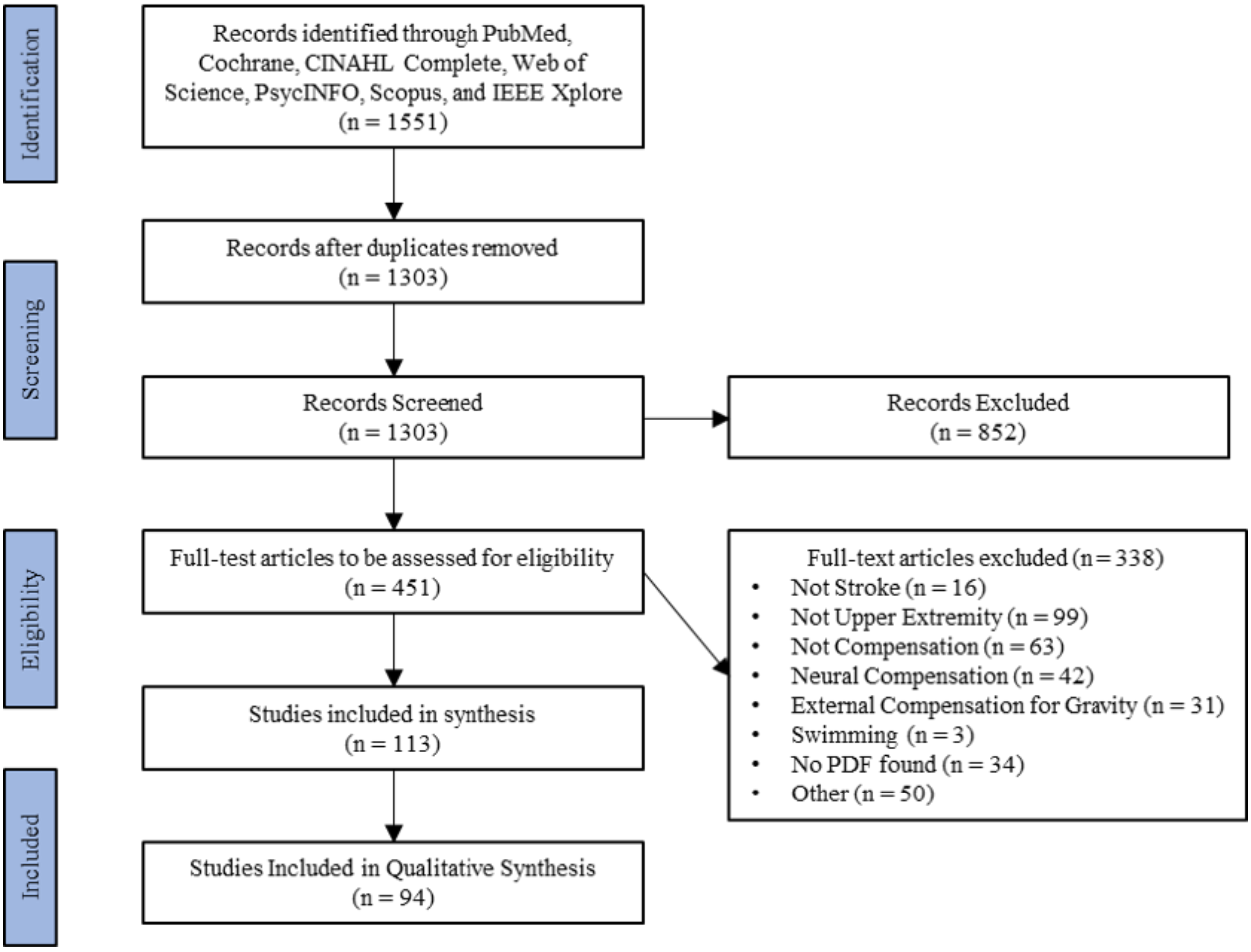


Figure 3.1: Flow chart depicting the article selection process

3.2.4 Charting data (Stage 4)

Data charting process: All of the articles which passed the screening process were included in the scoping review. A data extraction form was developed by the authors to extract and record any relevant information from the included source publications. The form was piloted with the research team; information was extracted from ten randomly selected articles reviewed by three investigators to ensure consistency. Refinements derived from this process resulted in a final version of the form (included in **Appendix A**) which was used to extract data from the 94 articles.

The information extracted included the kind of compensation defined in the article (e.g., non-use, behavioral compensation, motor compensation, and other compensatory behaviors not listed). If the compensation defined was a form of motor compensation, the segment that described as performing the compensatory behavior and the task being performed were also recorded. The specific, quantitative mechanism for defining compensation as well as the values or thresholds, units, and instruments used to measure compensation were recorded. Finally, participant demographics and trial design were recorded in the data extraction form.

Data items: For publications that met the inclusion criteria, the following variables were extracted: title, authors, journal, year, volume, pages, study type, compensation definition, and metrics used to describe compensation. For articles that investigated an intervention/treatment to reduce or prevent compensatory motor behavior, the following variables were extracted: number of subjects n, details on the specific treatments (e.g., dose, intensity, devices used), outcome measures, and primary study findings.

Risk of bias was determined for each study based on the following criteria: if a representative sample was included, subject randomization, intent-to-treat analysis, and blinded assessor. More than two negative scores on the criteria indicated a high risk of bias.

3.2.5 Collating and summarizing the results (Stage 5)

Synthesis of results: The investigators extracted the above information from each of the articles. First, it was validated a final time that the article focused on post-stroke, upper extremity compensation, and the general definition of compensation listed in the article

was recorded in the form. Then, the specific type of compensation was recorded, including the segment used for compensating and the task being performed when the compensatory behavior was measured. Then, if the compensatory movement was specifically quantified, the method or formula for quantifying the compensation, along with all reported numerical values and units of measurement, were recorded. Finally, the setting and device used to capture the compensatory behavior were recorded. The investigators then compared the results of the data extraction to validate that the reported results were accurate and addressed any inconsistencies by repeating the review of the articles until a consensus was reached. For the articles that investigated an intervention or treatment to reduce or prevent compensatory motor behavior, more demographics and statistics were recorded by the investigators.

The analysis of the articles showed that compensation is often specifically defined in the context of a segment being used to compensate and the task that is being performed while compensation is observed. However, the review also revealed that these specific definitions or descriptions of compensation are not always quantified. Finally, it was determined that the vocabulary used to describe compensation in the literature is inconsistent.

3.3 Results

3.3.1 Overview of included publications

The publication of articles describing compensation is a relatively recent phenomenon. Figure 3.2 shows the years in which the articles were published, indicating that of the 94 included articles nearly 70% were written in the last decade. The types of compensation described can be further divided into four categories: non-use, behavioral compensation, motor compensation, and *other* compensatory behaviors not listed. Motor compensation was most commonly reported, followed by non-use, behavioral, and other. Of the articles analyzed, only 72% specifically defined compensation.

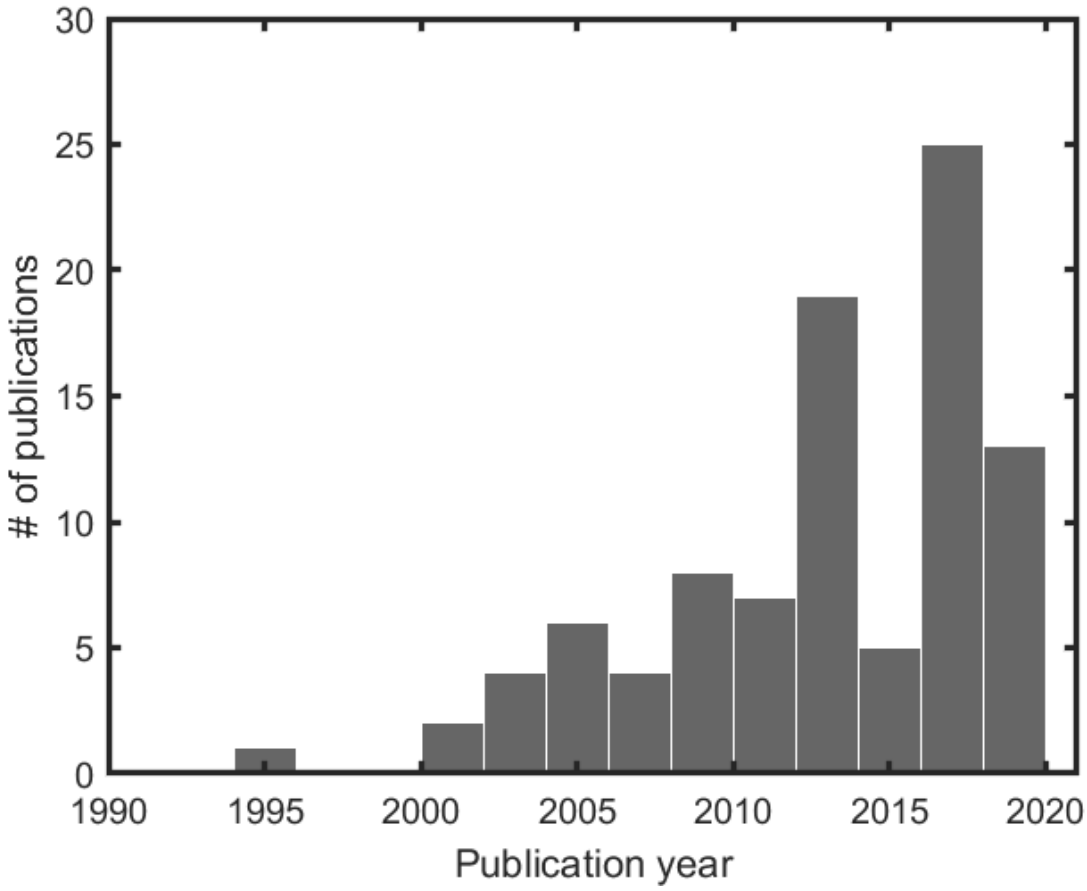


Figure 3.2: Publication years of selected articles

3.3.2 Explicit definitions of UE motor compensation

Levin's description of compensation [137] is widely accepted as the general definition for compensation during motor task performance. The article describes compensation as any way a person post-neurodegenerative disorder performs a task that is different than the way the task would have been performed before the onset of the neurodegenerative disorder. However, the initial goal of this analysis was to extract the specific definitions of motor compensation. Our analyses revealed certain generalities; in the bulk of the reviewed literature, compensation is not a term generally applied to describe a body-level impairment or capability; instead, compensation is almost always defined in the context of a segment or groups of segments, relative to a task being performed, or a combination of both a segment and a task. The majority of studies associated compensation with inappropriate recruitment of the trunk, followed by coupling of the shoulder or elbow joints during ADLs.

Segment-specificity of definitions

Of the articles focused on motor compensation, 84% specified the segment used to perform the compensatory behavior. The bar plot of Figure 3.3 depicts segments specifically identified during the performance of compensatory behaviors. The focus is primarily on truncal compensation with 72% of articles describing some compensatory engagement of the trunk. The shoulder and elbow are the next most described segments, followed by wrist and then fingers. Articles describing more than one segment engaged in the compensatory behavior show up in multiple columns.

Task-specificity of definitions

Compensatory behaviors were often defined in the context of a specific study task. Eighty-four percent of articles described the task used to elicit compensatory behaviors. The pie chart in Figure 3.3 depicts the types of tasks used for this purpose. The vast majority were reaches or reach-to-grasp tasks, followed by ADLs and study-specific tasks.

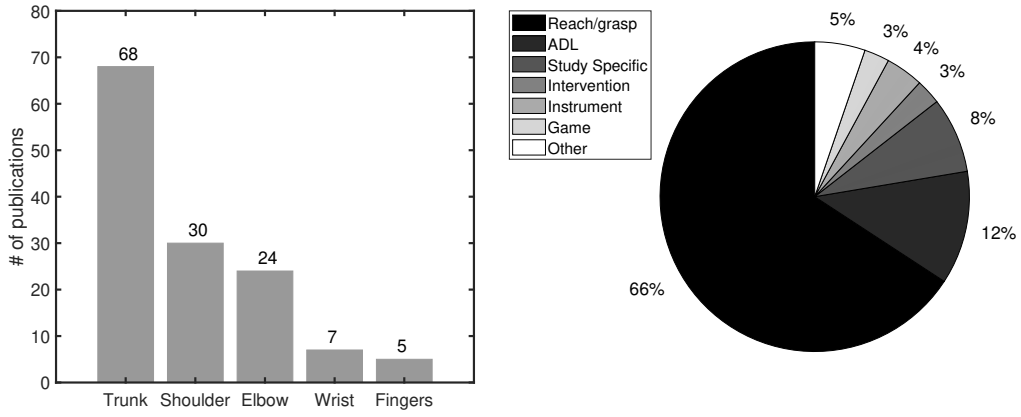


Figure 3.3: Segments (left) and tasks (right) for which compensatory behaviors are defined

3.3.3 Quantitative measures of UE motor compensation

Included publications can generally be grouped into separate classes based on the level of specificity of the compensation definition **Table 3.1**. These categories, arranged according to increasing specificity, include:

- Group 1: compensation is not described
- Group 2: compensatory motion is described, but without numerical specificity
- Group 3: compensatory motion is described, along with numerical values, but values are not explicitly related by authors to compensation
- Group 4: compensatory motion is described, along with numerical values, but any value > 0 is considered
- Group 5: compensatory motion is described, along with numerical values linked explicitly to compensation.

Table 3.2 depicts the terminology used to describe compensatory motion. The middle column contains the relevant segment; the third column contains the words that followed the segment, and the first column contains the terms that preceded the segment (e.g., *Excessive* **Trunk** *Compensation*, 16.67% of papers, $\rightarrow$ Column I, Column II, Column III).

3.4 Discussion

3.4.1 Explicit definitions of UE motor compensation

The purpose of this article was to determine quantitative descriptions of motor compensation. Initially, we expected that there would be a definition, a set of definitions, or a library of terms that could be used to specifically define the scope of compensatory behaviors. What we determined was the compensatory behaviors used and defined are dependent on both the task being performed and the segment being used to perform the task.

Segments associated with compensation: Truncal compensation was characterized most often in the papers which described compensatory behaviors. The engagement of

Table 3.1: Organization of publications according to increasing specificity of compensation definition

Group	Specificity	# of publications
1	Not described	11
2	Non-numeric	22
3	Numeric, unlinked	18
4	Numeric, linked, binary	15
5	Numeric, linked, granular	24

Table 3.2: Compensatory movements are specifically qualified a number of different ways across the papers we analyzed. This table shows the taxonomies that have been used to qualify compensation for each segment. The table is meant to be read [**Prefix**] **Segment** [**Suffix**], where the percentage following the suffix (or prefix if there is no suffix) is the percentage of papers which used this taxonomy to define a specific compensatory behavior. In instances where no prefix or suffix was used in the paper, that section of the table was left blank. In some instances, multiple prefixes preceded a single suffix, and in these cases, the percentages listed by the prefixes indicate the percentage of papers which used the listed prefix along with the listed suffix. The discrepancy between percentages listed for prefixes and suffixes in these cases is accounted for by the number of papers which used the suffix with no prefix.

Prefix	Segment	Suffix
	Trunk	-Arm Coordination (3.85%)
		-Assisted Reaching (2.56%)
		-Hand Distance (1.28%)
		Angle (5.13%)
		Angular Displacement (1.28%)
		Anterior Displacement (6.41%)
		Axial Rotation (1.28%)
Excessive (1.28%) Increased (3.85%) Reduced (1.28%)		Compensation (16.67%)
Excessive (1.28%)		Compensation Movements (8.97%)
Excessive		Compensatory Mechanisms (1.28%)
		Compensatory Offsets (1.28%)
		Compensatory Strategy (5.13%)
		Contribution (7.69%)
		Control (1.28%)
Arm- (2.56%) Compensatory Arm- (1.28%)		Coordination (2.56%)
Anterior (11.54%) Compensatory (5.13%) Excessive (6.41%) Forward (2.56%) Increased (3.85%) Sagittal (2.56%)		Displacement (39.74%)
		Elevation (1.28%)
		Excursion (2.56%)

Continued on next page

Table 3.2 – Continued

Prefix	Segment	Suffix
Anterior (1.28%) Compensatory (3.85%) Increased (2.56%) Δ (1.28%)	Trunk	Flexion (19.23%)
		Forward Bending (2.56%)
		Forward Path Length (1.28%)
		Forward Translation (1.28%)
Lateral (1.28%)		Inclination (1.28%)
		Involvement (1.28%)
		Lateral Flexion (2.56%)
		Lateral Path Length (1.28%)
Forward		Lean (1.28%)
		Leaning (1.28%)
		Medial Translation (1.28%)
Compensatory (5.13%) Excessive (6.41%) Increased (1.28%) Lateral (1.28%)		Motion (17.95%)
Additional (1.28%) Amount of (1.28%) Changes In (1.28%) Compensatory (29.49%) Excessive (20.51%) Forward (5.13%) Increased (3.85%) Reduced (1.28%)		Motion (42.31%)
Forward		Offsets (1.28%)
Arm-		Position (1.28%)
Increased (1.28%) More (1.28%)		Recruitment (11.54%)
Axial (1.28%) Compensatory (1.28%) Forward (1.28%) Lateral (1.28%)		Rotation (28.21%)
		Rotation Path Length (1.28%)
		Saggital Displacement (1.28%)

Continued on next page

Table 3.2 – Continued

Prefix	Segment	Suffix
	Trunk	Saggital Movement (1.28%)
External (1.28%)		Stabilization (1.28%)
		Tangential Velocity (2.56%)
		Trajectories (1.28%)
Compensatory (1.28%)		Use
Excessive (1.28%)		
Increased (1.28%)		
Compensatory Movement of the (6.41%)		
Excessive Flexion of the (1.28%)		
Excessive Rotation of the (2.56%)		
Increased Movement of the (1.28%)		
Involvement of the (1.28%)		
Involving the (1.28%)		
Recruitment of the (2.56%)		
Rotation of the (1.28%)		
Use of the (1.28%)		
Δ (2.56%)		
	Shoulder	-Elbow Interjoint Coordination (2.56%)
		-Elbow Coordination (1.28%)
		-Elbow Coupling (1.28%)
		-Elbow Index (1.28%)
		-Elbow Joint Motion (1.28%)
Degree of (1.28%)		Abduction (20.51%)
Excessive (1.28%)		
Increased (1.28%)		
Peak (1.28%)		
Synchronous (1.28%)		
		Abduction Angle (1.28%)
		Active Motion (1.28%)
		Adduction (8.97%)
Range of (1.28%)		Angle (3.85%)
		Angular Displacement (1.28%)
		Compensation (2.56%)
Disrupted Elbow-		Coordination (1.28%)
		Displacement (1.28%)
		Elevation (7.69%)
		Extension (6.41%)
		External Rotation (5.13%)

Continued on next page

Table 3.2 – Continued

Prefix	Segment	Suffix
Active (2.56%) Maximum (1.28%)	Shoulder	Flexion (20.51%)
		Flexion Angle (1.28%)
		Flexion Torque (1.28%)
		Girdle Movement (1.28%)
		Horizontal Abduction (3.85%)
		Horizontal Adduction (8.97%)
		Internal Rotation (2.56%)
		Inward Rotation (1.28%)
		Joint Angle (1.28%)
		Joint Contribution (1.28%)
		Joint Correlation (1.28%)
Under-use of		Joints (1.28%)
Deficits in		Kinematic Adaptability (1.28%)
		Moment (1.28%)
Maximal Velocity of		Motion (1.28%)
Amount of (1.28%) Compensatory (2.56%) Limited (1.28%) Range of (1.28%) Temporal Coordination of (1.28%)		Movement (8.97%)
		Protraction (1.28%)
		Range of Motion (1.28%)
		Rotation (5.13%)
		Scapular Elevation (1.28%)
		Subluxation (2.56%)
		Upward Compensation (1.28%)
Maximal		Velocity (1.28%)
Compensatory Movement of the (1.28%)		
Rotation of the (1.28%)		
	Elbow	-Shoulder Coordination (1.28%)
		-Shoulder Interjoint Coordination (2.56%)
		-Shoulder Joint Motion (1.28%)
		-Shoulder Relationship (1.28%)
		Active Motion (1.28%)
		Angle (3.85%)
		Angular Excursion (1.28%)
Shoulder-		Coordination (2.56%)
Shoulder-		Coupling (1.28%)
		Displacement (1.28%)

Continued on next page

Table 3.2 – Continued

Prefix	Segment	Suffix
	Elbow	Excursion (1.28%)
Maximum (2.56%)		Extension (23.08%)
		Extension Amplitude (1.28%)
		Extension Range (1.28%)
		Extension Velocity (1.28%)
Amount of (1.28%) More (1.28%)		Flexion (14.1%)
		Flexion Torque (1.28%)
		Force (1.28%)
Shoulder-		Index (1.28%)
Shoulder-		Interjoint Coordination (2.56%)
		Joint Angle (1.28%)
		Joint Contribution (1.28%)
Under-use of		Joints (1.28%)
Deficits in		Kinematic Adaptability (1.28%)
		Moment (1.28%)
Amount of (1.28%) Temporal Coordination of (1.28%)		Movement (12.82%)
		Pronation (1.28%)
		Range-of-Motion (1.28%)
Velocity of		Rotation (1.28%)
		Supination (1.28%)
	Wrist	Angle (1.28%)
Impaired		Control (1.28%)
		Coordination (1.28%)
		Extension (1.28%)
		Extension Angle (1.28%)
		Extension Strength (1.28%)
		Flexion (1.28%)
		Movement (2.56%)
		Peak Velocity (2.56%)
Maximal (1.28%) Peak Horizontal (1.28%)		Velocity (1.28%)
Magnitude of (1.28%) Reduced (1.28%)	Finger	Abduction
Between-		Covariances (1.28%)
Impaired Coordination of		Forces (1.28%)
		Joint Angular Excursions (1.28%)

Continued on next page

Table 3.2 – Continued

Prefix	Segment	Suffix
	Finger	Joint Coordination (1.28%)
		Joint Excursions (1.28%)
Angular Excursion of		Joints (1.28%)
		Movement (1.28%)
		Range of Motion (1.28%)
Impairment in the (1.28%)		

the trunk to complete a task, such as using forward lean to supplement the execution of a reach, is easy to measure, address, and reverse as these movements are distinctly different from control performance of the same task in both an observable and quantifiable sense. Compensation using the shoulder and elbow were described less than half as often as compensation using the trunk. While these segments do demonstrate some individual compensatory behaviors, compensation of the shoulder or elbow can be more difficult to observe or measure since it often includes individual compensatory strategies combined with decreased synergy and coordination between the two segments. Lastly, only a small number of papers analyzed described compensation of the wrist or fingers as anything more than weakness or inappropriate orientation for a given task. These task-specific definitions indicate that there are fewer general definitions of compensation available for these segments than the other three.

Constraint and Compensation: For the trunk, shoulder, and elbow segments, compensation is frequently quantified as more or less engagement of the segment to complete a task. The threshold for what is considered compensatory is set by the level to which the segment would have been engaged before the onset of stroke. This is exemplified in the papers which describe the increased trunk movement, decreased should flexion, and decreased elbow extension that persons post-stroke use to complete a reaching task. The understanding that is expected from these definitions of compensation is that some engagement of these segments is anticipated during the execution of such tasks. There is, however, a binary method which is occasionally used for describing compensation where the "threshold" is set to define any engagement of a given segment during the completion of a task as compensatory. This binary of description of compensation is often used in papers which seek to prohibit use of the segment, such as those which restrain the trunk during reaching tasks or constrain the less affected limb to reverse non-use of the more affected limb. These binary definitions of compensation are used in study design as often as they are because providing a clear cut-off point makes it straightforward to decide whether the compensatory behavior is occurring and if intervention is necessary. This allows researchers to more easily measure and develop therapies to reverse or prevent compensation defined in this way.

Compensation and the Selected Task: Analysis of the papers also revealed that compensation is most often identified in the context of a task being performed. This indicates that description and quantification of compensatory behavior is task specific, not generally applicable. The task during which compensation was most analyzed was a prototyped reach or reach-to-grasp task. Reaching tasks are likely to invoke forward truncal compensation which, as mentioned, is easy to measure, address, and reverse. However, although reaching can be an integral part of functional task performance in the ambient setting, prototyped tasks such as those that are performed in the papers analyzed are not representative of the types of behaviors for which compensatory strategies are typically employed. As such, some studies asked participants to perform tasks meant to simulate those performed in the ambient setting. Other papers developed games or study-specific tasks which occasionally use specialized instruments to evaluate and reverse compensatory behaviors. The reviewed articles reveal that compensation cannot be specifically quantified without first describing the relevant behavior and/or the segment(s) used.

3.4.2 Identify specific, quantitative measures of UE motor compensation assigned to each explicit definition (if given), and their prevalence of use in published literature.

Quantitative Descriptions of Compensation: The secondary goal of this article was to develop an agreed upon quantitative measure for upper extremity compensation. Whether attempting to measure compensation in the lab or ambient setting, the capability to quantify compensation autonomously, without the aid of a therapist or clinician, can be extremely useful for providing a clinician with information that can be used to design or tune individualized rehabilitation programs for patients. As previously discussed, compensation is most easily and often defined in the context of a segment being used and a task being performed, and using those definitions, the papers were divided into groups based on the granularity that was used to define the compensatory behavior.

Compensation, while generally well defined, is not well quantified. Despite this, nearly a quarter of the papers discussed methods for measuring compensation or interventions

to decrease compensation without quantifying the compensation that was being addressed. Other papers binarily quantified compensation (segment was used or not used during a task) or defined compensation in the context of an intervention (segment was used less during a given task by a control group or after an intervention). Part of the purpose of publication is to present replicable results that can be used as a foundation for future research. The papers which were placed in the top group provided definitions of compensation in the context of a segment used, a task being performed, and quantified values of the described compensation. These definitions can be extracted by unaffiliated researchers and used in novel sensing methods or therapies. Since only approximately a quarter of the papers provide specific, quantitative definitions of compensation, more effort needs to be made in the future to narrate consistent definitions of compensation.

3.4.3 Determine any existing taxonomies for such compensation.

Analysis of the articles revealed that compensation is designated in a variety of ways in the literature. While it is important to evaluate the specific quantitative ways that people are describing compensation, if the nomenclature used for compensation too varied, the disparate descriptions cannot be combined to develop a cohesive definition. With regards to compensation, which should be well defined or described but is not, it is inappropriate to use an array of terms since consistencies between definitions can be loss because of inconsistencies between terms. The taxonomies of the trunk which are demonstrated in **Table ??** often overlap in the type of compensatory movement that is being referenced despite the difference in terminology. In the future, the best practice would be to stick to a limited, but widely accepted, nomenclature for describing compensation or compensatory behaviors.

3.5 Implications for Our Approach

This scoping review was conducted to determine the existing metrics for quantifying compensation in the state-of-the-art-literature. We found that compensation is often described in the context of a segment used to compensation and a task being performed when compensation is measured. However, beyond reaching, there were no prototyped

motions for which compensation was quantitatively described. Therefore, there is a need to classify motion generally before compensation can be specifically quantified. In later chapters, we will look at quantifying known maladaptive movement strategies to measure whether significant differences exist between tasks being performed. If significant differences emerge, then the quantification of such movement strategies can later be used as a proxy for compensatory behavior.

We also developed a tiered list of the ways that compensation is described in the literature and, in so doing, a standard for describing compensation that authors can strive for in future work. Finally, we demonstrated the various ways that compensation is described for each segment. This review reveals the need for consistency in the methods used to describe and quantify compensatory behavior as well as the dearth of specific, quantifiable definitions of compensation. Future work in the field of sensing and reversing upper extremity compensatory movement might consider the standards suggested by this paper and strive for a more cohesive attempt at quantifying such behavior.

Chapter 4

Novel Method to Quantify Interlimb Coordination in Persons with Hemiparesis Post-Stroke

4.1 Introduction

In **Chapter 3**, we determined that compensation is primarily described in the context of a task being performed and a segment being used to perform a task. Coordination of the upper limbs, which may be impaired post-stroke [216], is important for many daily life skills, and the amount of coordination needed is dependent on the task being performed [80]. In this chapter, we use minimally intrusive sensing techniques to quantify interlimb coordination. We then use the derived values to prove significant differences between task type and correlation between the derived values and clinical measures.

Finely-tuned upper extremity (UE) intra- and interlimb coordination is controlled through intact neural coupling [8], which requires timing of movements and sequential, rhythmic use of limb segments on one or both sides of the body [237]. Coordination of the upper limbs enables interaction with the environment and the performance of goal-oriented tasks such as activities of daily living (ADLs). The type of tasks performed range from unimanual with use of a single limb, to bimanual symmetric or the production of mirrored

UE movements, to bimanual asymmetric when different motion is exhibited in each limb. For persons with hemiparesis post-stroke, tasks requiring UE coordination can be difficult to execute due to limited strength, mobility, and motor control resulting in the execution of functional, yet compensatory movement patterns [216, 235, 119].

Compensatory strategies may include disuse of the more affected limb, increased trunk involvement during arm motion, asymmetry during bimanual tasks, and inefficient motion or atypical synergistic movements during task performance [138, 245, 294]. Although compensation promotes independence in everyday tasks and may be ideal for individuals with severe impairments, it can also impede recovery of intra- and interlimb coordination inherent in unimanual and bimanual performance [119, 137] as persistent disuse or the display of inefficient patterns may lead to reduced capability and cortical representation of the more affected limb [241, 177]. Determining the extent of coordination within and between the arms and how it changes with recovery and rehabilitation requires assessment measures sensitive to subtle changes in motion and task performance.

Most UE clinical assessments evaluate function of the paretic limb during unimanual tasks with limited emphasis on bimanual function [75, 283, 297]. An exception to this is the Assisting Hand Assessment (AHA) [93, 124], a tool originally designed to assess how effectively the more affected limb is used during bimanual tasks in children with unilateral UE dysfunction. The AHA has been recently adapted for use in adults post-stroke, i.e. the Adult AHA Stroke (Ad-AHA Stroke) [125]. However, as with other observation-based tools, the Ad-AHA Stroke may not be sensitive enough to detect small yet significant changes in motor behavior occurring with natural recovery or rehabilitation. A highly sensitive, objective measure requiring minimal equipment is needed to quantify intra- and interlimb coordination for different tasks in various settings.

The ability to perform coordinated UE tasks is an important rehabilitation outcome that is often neglected according to patients and clinicians [21, 20]. Existing literature on interlimb coordination has largely focused on the evaluation of gait [57, 231, 247], or the use of tethered marker–based techniques unsuitable to ambient settings [216, 292]. As reinforcement of compensatory behaviors is thought to primarily occur in ambient settings when an individual is unmonitored by a health care specialist [177], existing research approaches that evaluate

limb coordination using kinematic metrics are inappropriate for ambient monitoring given the requirement that users are stationary or tethered to measurement equipment. This has motivated the development of various minimally intrusive sensor–based approaches to monitor motion in individuals with motor dysfunction [257, 243, 199].

Inertial measurement units (IMU), are body-worn sensors that monitor and transmit changes in movement during the execution of everyday tasks [88]. IMU sensors have been used with individuals post-stroke and other neurological conditions to capture the quality and quantity of motion during typical and atypical motor behaviors [88, 253]. These IMU sensors have also been used with the stroke population to predict impairment scores, functional scores, and activity performance [298, 197, 59]. These sensors can detect small changes in movement patterns, thus should be able to differentiate in a quantitative manner between typical and atypical motor behavior. A challenge in using IMU sensors is that they produce derived, differential motion measures, such as linear acceleration and angular rate of change. Therefore, unlike traditional marker-based motion capture systems, raw data cannot be easily used to directly reconstruct changes in limb position. Instead, IMU data requires custom signal and data processing techniques to produce clinically relevant metrics [268].

The goal of the current approach is to develop novel sensor based measures of UE interlimb coordination as a potential biomarker for motor recovery. Development of an accurate yet sensitive system using IMU data to identify distinct features of UE intra- and interlimb coordination is a sequential process. We will develop a novel algorithm for the differentiation of unimanual, bimanual asymmetric, and bimanual symmetric gestures using inertial sensor data. This algorithm will show that individuals post–stroke demonstrate quantifiable kinematic differences in the performance of these gestures when compared to healthy individuals [113].

The first aim of this chapter is to describe this novel algorithm and provide an exploratory evaluation of the ability of a sensor-derived motion parameter to differentiate between UE performance in healthy controls and individuals post stroke. The second aim is to evaluate the ability of this motion parameter to distinguish between UE task type (unimanual, bimanual symmetric, and bimanual asymmetric tasks) in healthy controls. If successful, the completion of these two aims will result in the validation of an algorithm capable of

distinguishing between types of interlimb coordination based on data from inertial sensors. The final aim is to validate findings from a sensor-derived motion parameter against clinical measures commonly used to assess performance in persons post-stroke, including the UE Fugl-Meyer Assessment (UE-FMA) and the Ad-AHA Stroke Assessment. If the sensor derived data are sensitive to differences in groups (controls and post-stroke) and task and if the results are validated against clinical measures, data from these participants can be used in later chapters (**Chapter(s) 5, 6**) to train machine learning models to autonomously classify group and task type.

The methodology used to develop our algorithm is described in Section 4.2. The results of our study are presented in Section 4.3, and interpretation and implications of our results are presented in Section 4.4.

4.2 Methods

4.2.1 Participants

Individuals post stroke and healthy-age matched controls were recruited from two clinical sites: Columbia University Irving Medical Center/Teachers College, Columbia University and Chapman University. Inclusion criteria for participants post stroke were: (1) $>$ 1 year, post stroke; (2) ability to isolate elbow and shoulder motion in one arm; and (3) ability to perform a gross grasping task, e.g. 3-jaw chuck or lateral pinch with both hands. Exclusion criteria included: (1) atypical synergy patterns involving concurrent motion in 2 UE joints; (2) joint contractures $> 20°$ at either elbow or $> 45°$ in either shoulder; and (3) known allergies to tape or other skin sensitivities. Potential participants were recruited using flyers and referrals from existing databases. This study was approved by the Institutional Review Boards at Columbia University Irving Medical Center, Teachers College, Columbia University and Chapman University. All participants provided written informed consent.

4.2.2 Clinical Characteristics and Demographics

All participants were assessed with the Edinburgh Inventory to determine handedness [192], the Manual Ability Measure-36 (MAM-36) to assess hand function [40], and the Jamar® Dynamometer and BL Engineering Pinch Gauge to asses grip and pinch strength [222]. The MAM-36 rates 36 everyday tasks based on self-reported manual ability with the total score ranging from 0–144; higher scores indicate better-perceived manual ability. Grip, lateral pinch, and palmar pinch strength are reported as the mean of three trials. Persons post-stroke were also assessed using the UE-FMA to establish motor impairment based on a maximum score of 66 with a higher score indicating better motor function [75, 144, 194]. Participant characteristics are shown in **Table 4.1**.

4.2.3 Experimental Procedure

In this cross-sectional study, participants attended a single session during which they completed six UE tasks and the Ad-AHA Stroke [125] while seated and wearing five APDM Opal wearable sensors (Portland, OR). Performance was videotaped to obtain accurate start and stop times for all tasks. The sensors were secured with adjustable straps on each wrist, each upper arm, and the sternum (**Fig. 4.1A**). Each sensor recorded tri-axial acceleration, angular rate of change, and magnetic field strength at 128 Hz. Prior to collecting data during the UE tasks and Ad-AHA Stroke, one sensor was shaken in a rhythmic pattern to facilitate post–hoc synchronization of inertial and video data. The six UE tasks (see sample **Fig. 4.1B**) were counterbalanced for order and completed twice by each participant. Tasks included two unimanual (U) tasks (reaching for a bottle and reaching across midline for a spoon); two bimanual symmetric (BS) tasks (folding a towel and donning a hat); and two bimanual asymmetric (BA) tasks (unscrewing a bottle lid and stirring marbles in a bowl). In principle, there should be no kinematic coupling between limbs for unimanual tasks, a high level of coupling for bimanual symmetric tasks, and moderate coupling for bimanual asymmetric tasks.

Paretic limb performance in persons post-stroke was further assessed using the Ad-AHA Stroke [125], which required participants to unwrap and wrap a present (**Fig. 4.1C**).

Table 4.1: Participant Characteristics

ID	Age (yrs)	Post-Stroke (yrs)	Gender	Affected Side	EI	Handedness	FM	MAM	Grip (kg) U/A, D/ND	Lateral Pinch (kg) U/A, D/ND	Palmar Pinch (kg) U/A, D/ND
1	67	4	F	L	80	R	29	107	28/4	7.6/2.5	7.3/0.0
2	60	11	M	L	100	R	9	112	32/5	4.0/0.0	6.0/1.0
3	52	10	F	L	100	R	35	122	29/12	4.0/2.0	5.3/2.5
4	64	13	F	R	-90	L	29	130	30/2	4.0/0.5	4.0/0.8
5	67	11	M	R	-60	L	25	110	25/8	4.0/0.0	3.0/0.0
6	25	11	M	L	70	R	30	134	48/18	6.0/0.0	4.0/1.0
7	64	8	F	R	-100	L	32	112	25/4	6.0/0.0	4.0/0.5
8	41	4	F	R	-70	L	27	123	25/8	4.0/4.0	5.0/0.0
9	50	3	M	R	-80	L	26	121	45/15	7.5/2.0	7.0/2.5
10	43	10	M	R	-100	L	31	114	35/8	8.0/2.0	5.0/1.0
11	52	2	M	L	100	L	21	33	43/2	11.5/1.5	10.0/0.0
12	56	3	M	R	27	Am	64	121	37/26	9.6/7.6	6.0/6.3
13	66	8	M	R	-62	L	31	130	33/8	7.9/2.7	5.8/1.3
14	67	5	M	L	90	R	35	107	32/4	9.5/2.8	5.0/0.5
15	59	2	M	R	80	R	45	123	30/18	9.0/5.8	6.8/6.0
16	40	6	M	L	78	R	44	103	40/8	8.0/6.0	6.1/1.1
17	50	16	M	R	100	R	64	130	40/35	8.4/9.1	7.0/5.8
18	42	8	F	L	100	R	42	105	23/1	6.7/2.0	4.2/0.3
19	46	11	M	L	68	R	28	99	36/6	8.5/2.3	7.0/1.3
20	58	8	F	L	100	R	35	114	36/6	5.83/1.3	4.0/1.2
Controls n=20	52.3 (12.5)	NA	5M/15F	NA	75 (37)	18R/2Am	NA	143 (1.5)	32.3 (7.4)/ 29.3 (7.4)	6.9 (2.4)/ 6.5 (2.5)	5.6 (2.2)/ 5.2 (1.7)

ID = identification; yrs = years; F = female; M = male; EI = Edinburgh Inventory; R= right; L = Left; Am = Ambidextrous; FM = Fugl-Meyer Motor Assessment; MAM = Manual Ability Measure; kg = kilograms; U = unaffected; A = affected; D = dominant; ND = non-dominant.

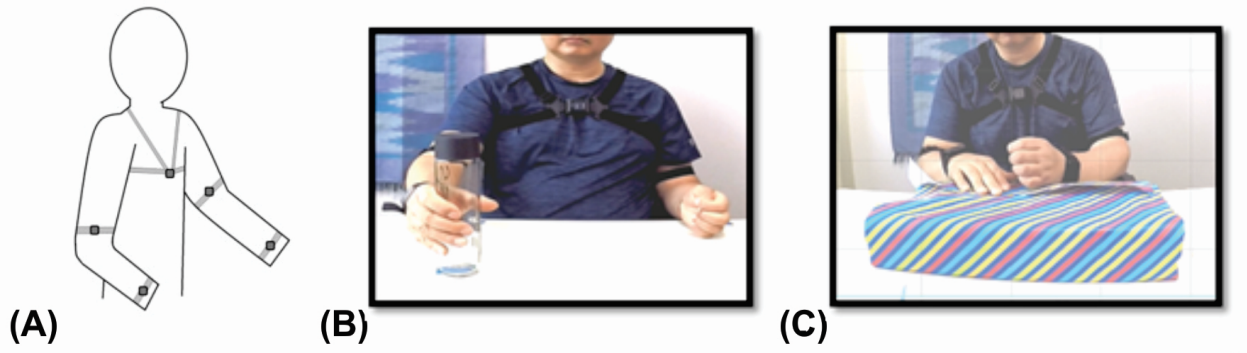


Figure 4.1: Subjects wore 5 sensors on the sternum, L/R wrist, and L/R upper arm (A) during performance of 6 different tasks performed twice (B), including reaching for a water bottle as shown here; and performance of the Adult Assisting Hand Assessment Stroke (AdAHA-Stroke), which involved unwrapping and wrapping a present (C).

Specifically, participants were informed that the aim of the task was to use both hands in the way that felt most natural and easiest. This would allow assessment of functional performance when both hands were used together. Scoring of the Ad-AHA Stroke from videotape was done on 19 components within five categories: general usage, arm use, grasp-release, fine-motor adjustment, and coordination. Sample components included items such as initiates use, stabilizes by grip and flow in bimanual task performance. A 1-4 category rating scale was used to score the more affected limb on all 19 components: 1) does not do; 2) ineffective; 3) somewhat effective; and 4) effective. The Ad-AHA Stroke has established validity and reliability in individuals post stroke [264]. All ratings were completed by an experienced and certified rater.

4.2.4 Data Processing

Use of IMU data requires custom signal and data processing techniques to produce clinically relevant metrics [268]. Raw sensor data were filtered using a 3rd order Butterworth bandpass filter [170] with 0.1Hz and 2Hz cutoff frequencies and then digitally de–trended to remove drift [268]. The adjusted data were then separated into twelve different segments corresponding to each task. Sample control and post–stroke data corresponding to the *Fold Towel* task is presented in Fig. 4.2.

4.2.5 Interlimb Coordination Measures

Investigation of gesture coupling and coordination typically involves the comparison of segments or anatomical landmarks using marker based measures that rely on kinematic (position) data relative to a global coordinate frame [230, 8, 85]. Quantifying coordination with inertial sensors is challenging because 1) inertial sensors utilize a local coordinate frame (versus a global coordinate frame), and 2) inertial sensors measure acceleration and angular rate of change (versus position and orientation). To account for this, we utilized three intuitively motivated comparison approaches evaluating peak to peak amplitude, time domain properties, and frequency domain properties.

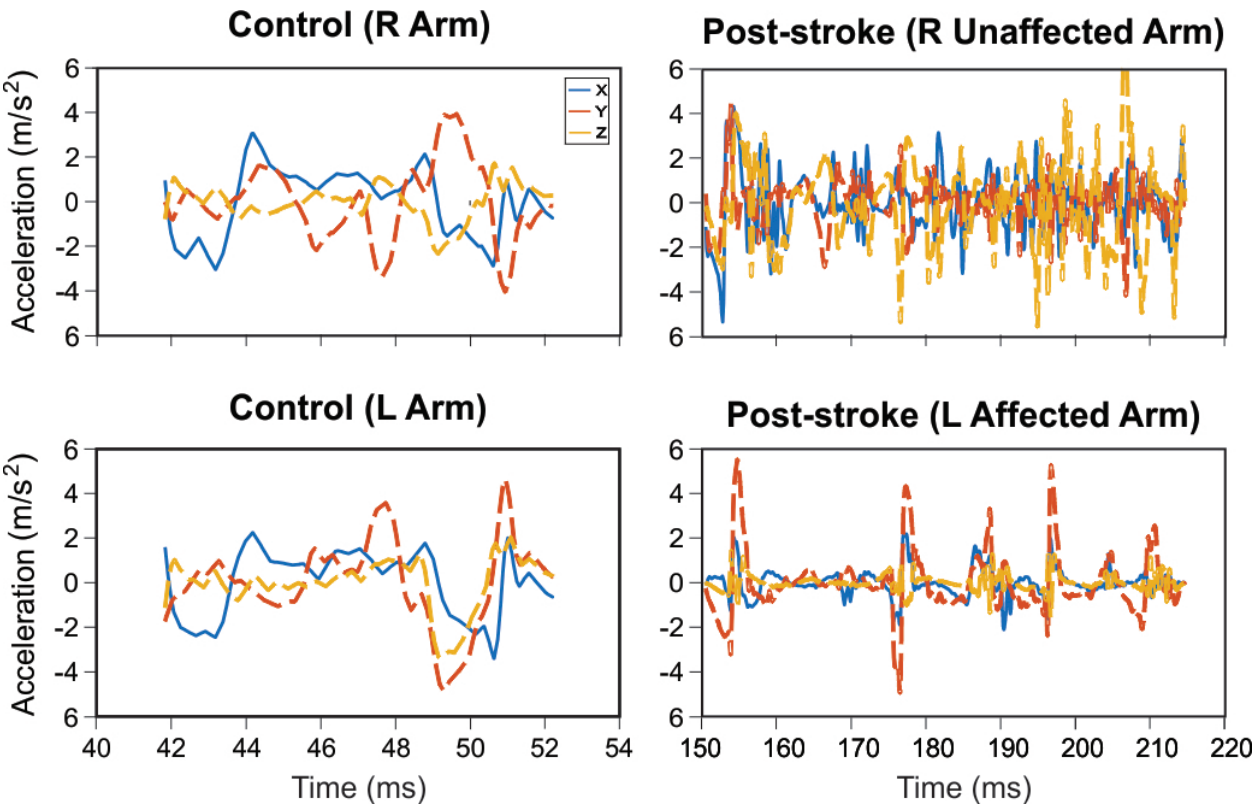


Figure 4.2: Raw acceleration data extracted from the left and right wrists of a control (left side plots) and persons post–stroke (right side plots) performing the task *Fold towel*.

Peak to Peak Amplitude

Peak to peak amplitude comparisons are based on the assumption that similar motions generate similar amounts of amplitude in the IMUs; yet, they do not account for signal morphology. The peak to peak amplitude of each axis for each sensor modality was determined, then normalized to the maximum value. The peak to peak amplitude for each sensor and each modality was then compared to the peak to peak amplitude of every other sensor for the same modality (e.g., all accelerometer data were compared to each other).

Time Domain

Time domain comparisons are more sensitive to morphology as they consist of point by point comparisons; yet, they may be overly sensitive to noise and small signal fluctuations. Spearman's correlation was used to compare each axis to every other axis for all sensors. Given the number of axes and sensors, the correlation coefficients were then compressed around a mean value. To minimize the associated loss of information, the correlation values were ranked from smallest to largest. These values were then multiplied by their rank, resulting in a weighted average such that high correlation values carried heavier weight than low values. The ranking was applied to a $3n \times 3n$ matrix (corresponding to the number of axes, 3, multiplied by the number of sensors, n).

Frequency Domain

Frequency domain measures have demonstrated validity in capturing the underlying structure of motion; however, they may be more difficult to evaluate intuitively [48, 269]. In the frequency domain, similarity was quantified using the magnitude–squared coherence (MSC) between each axis on every sensor for a given gesture. This measure provided an indication of frequencies for which the two signals had overlapping signal power, and was calculated using:

$$C_{xy} = \frac{|P_{xy}|^2}{P_{xx}P_{yy}} \tag{4.1}$$

where x and y represent two signals, and P_{ij} is the power spectral density of two signals i and j. To convert the resulting vector into a scalar, the area under each C_{xy} curve was

computed and normalized to produce a matrix of values between zero and one. Like the time domain similarity measure, each frequency domain similarity measure was assigned a rank and multiplied by that rank.

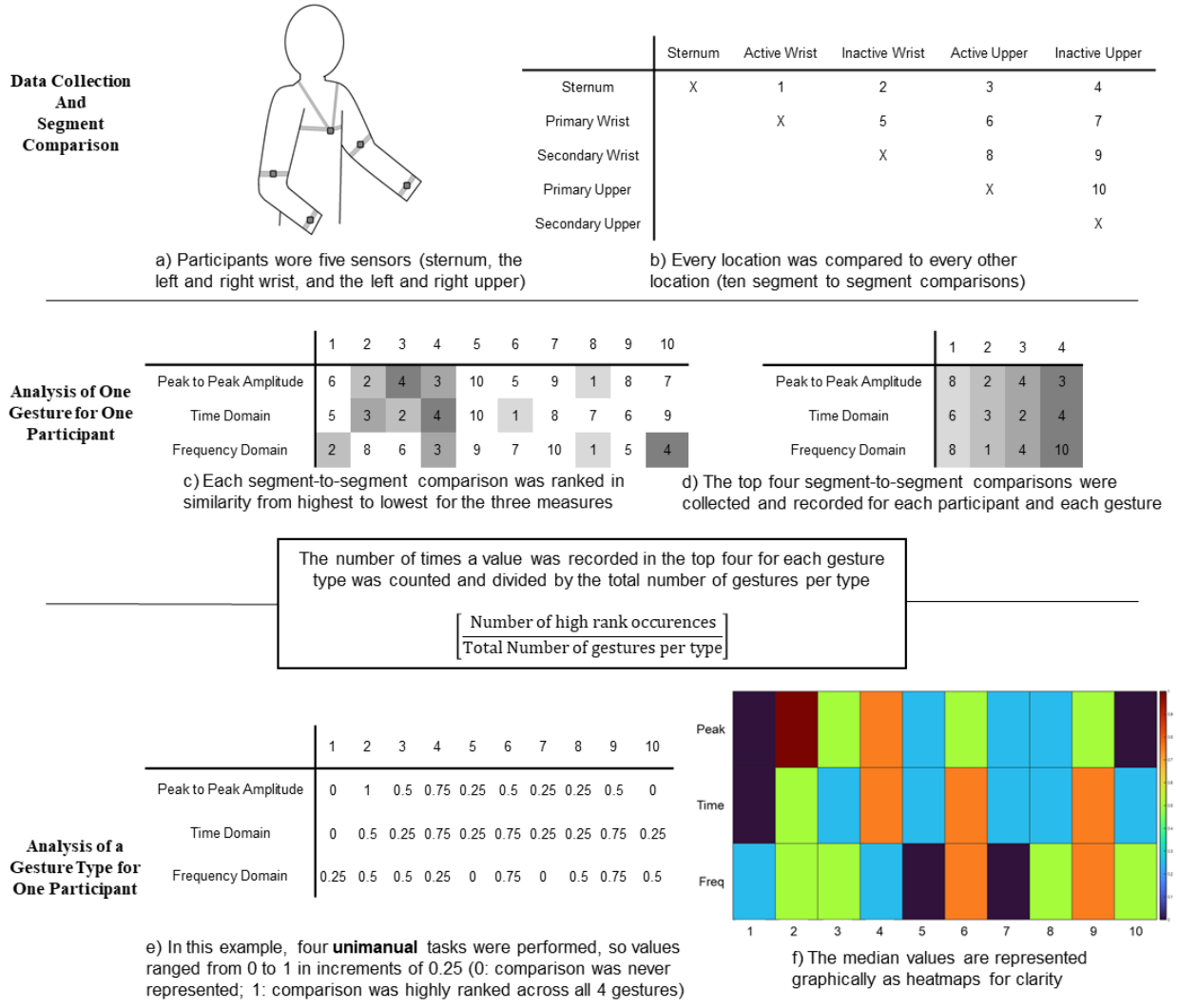
4.2.6 Interlimb Coordination Algorithm

Our novel interlimb coordination algorithm depicted in Fig. 4.3 uses all three similarity metrics with five IMUs used in the calculations (Fig. 4.3a). Each sensor was compared to each other sensor, resulting in ten comparisons (Fig. 4.3b). The wrist 1 and upper arm 1 sensors were classified as more active and the wrist 2 and upper arm 2 sensors as less active during each task. Specifically, we determined which was the leading limb for bimanual tasks and which was the moving limb for unimanual tasks based on performance of controls.

The ten sensor pairs were: a) sternum to wrist 1 (S-W1); b) sternum to wrist 2 (S-W2); c) sternum to upper arm 1 (S-U1); d) sternum to upper arm 2 (S-U2); e) wrist 1 to wrist 2 (W1-W2); f) wrist 1 to upper arm 1 (W1-U1); g) wrist 1 to upper arm 2 (W1-U2); h) wrist 2 to upper arm 1 (W2-U1); i) wrist 2 to upper arm 2 (W2-U2); and j) upper arm 1 to upper arm 2 (U1-U2).

To determine the similarity between the motion of any two segments (e.g., wrist 1 to upper arm), data from each modality (e.g., acceleration) was compared between sensors for three metrics: peak to peak amplitude, the correlation in time domain, and the frequency domain (spectral coherence). Thus, data from the x-, y-, and z- axes of one sensor was compared to the data from the x-, y-, and z- axes of another sensor using each of the three metrics. For the three modalities, this resulted in nine values each for relative amplitude, time domain, and spectral coherence [178]. The nine values for each modality were then normalized to the maximum value, and the resulting twenty-seven normalized values were averaged to create a single value for each comparison, titled the motion parameter.

The aforementioned analyses of task performance for each participant generated a single motion parameter or similarity value between each sensor pair for each of the three measures (i.e., peak-to-peak amplitude). The sequence for our algorithm is shown in **Fig. 4.4**. To assess intra and interlimb coupling the motion parameters for the ten sensor pairs were



	Sternum	Active Wrist	Inactive Wrist	Active Upper	Inactive Upper
Sternum	X	1	2	3	4
Primary Wrist		X	5	6	7
Secondary Wrist			X	8	9
Primary Upper				X	10
Secondary Upper					X

	1	2	3	4	5	6	7	8	9	10
Peak to Peak Amplitude	6	2	4	3	10	5	9	1	8	7
Time Domain	5	3	2	4	10	1	8	7	6	9
Frequency Domain	2	8	6	3	9	7	10	1	5	4

	1	2	3	4
Peak to Peak Amplitude	8	2	4	3
Time Domain	6	3	2	4
Frequency Domain	8	1	4	10

	1	2	3	4	5	6	7	8	9	10
Peak to Peak Amplitude	0	1	0.5	0.75	0.25	0.5	0.25	0.25	0.5	0
Time Domain	0	0.5	0.25	0.75	0.25	0.75	0.25	0.25	0.75	0.25
Frequency Domain	0.25	0.5	0.5	0.25	0	0.75	0	0.5	0.75	0.5

Figure 4.3: The interlimb coordination algorithm utilizes amplitude, time, and frequency data from all five body-worn sensors to create an array of numerical values for unimanual, bimanual asymmetric, and bimanual symmetric gestures.

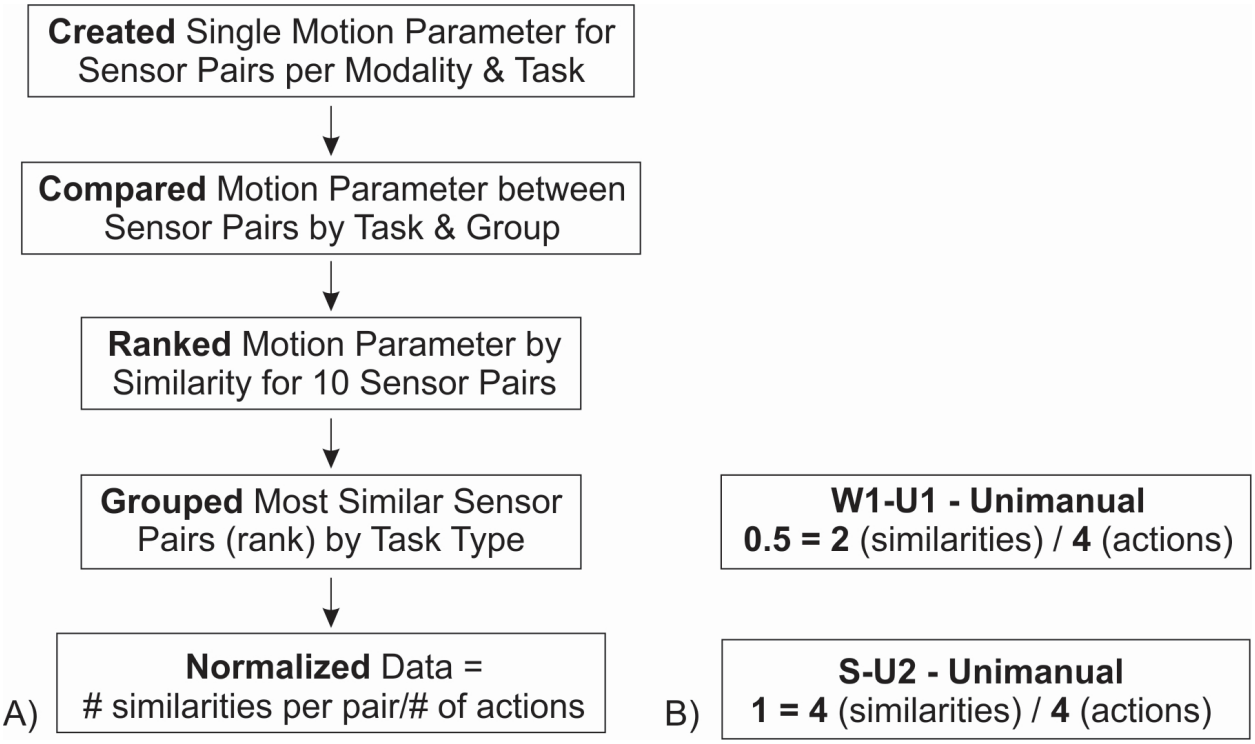


Figure 4.4: A) Sequence of algorithm construction and B) two examples.

ranked from highest to lowest in regard to the motion parameter (Fig. 4.3c). To normalize the data for task type, we divided the number of times each sensor pair had a high similarity value by the number of actions performed (Fig. 4.3d). To group these results within gesture type (e.g., unimanual gestures), we determined the frequency at which a comparison showed up as a high rank comparison. Because participants performed four different gestures for each gesture type (unimanual, bimanual asymmetric, and bimanual symmetric), this frequency could take on values of 0, $\frac{1}{4}$, $\frac{1}{2}$, $\frac{3}{4}$, or 1 (Fig. 4.3e). Finally, we visualized the similarity values using a heatmap (Fig. 4.3f). For each participant, the algorithm outputs three data sets (similar to that of Fig. 4.3f) corresponding to unimanual, bimanual asymmetric, and bimanual symmetric data.

Our algorithm was based on task type and performance by controls. For example, each participant performed four unimanual tasks. We would expect a control to have a high similarity value between the sternum and inactive upper arm (S-U2) for all four iterations of unimanual tasks since both body parts are apt to move very little during task performance. This sensor comparison would result in a similarity score of 4/4 or 1. Conversely, since data from the wrist and upper arm sensors of the active limb (W1-U1) of a control may only be fairly similar (same motion but different amplitude), that sensor pair might only demonstrate similarity for two of the four task iterations resulting in a similarity score of 2/4 or 0.5. Thus, the top ranked sensor pairs per task type with the highest frequency often received scores closer to "one," while sensor pairs that were low ranking often received scores closer to "zero." The ranking was performed for each task type and modality, and the similarity values for each participant post-stroke were compared to an age-matched control.

In persons post-stroke, we extracted a use ratio during performance of the Ad-AHA Stroke by comparing the wrist sensor data from the paretic arm against the non-paretic arm (paretic/non-paretic arm). This use ratio was then compared to scores obtained for the UE-FMA as well as the Ad-AHA Stroke. For each sensor modality (acceleration, angular rate of change, and orientation), the total area under the raw data curve (integral) for the paretic arm was normalized by the total area under the curve (integral) for the non-paretic arm. Thus, the use ratio for each modality was compared again the UE-FMA and AdAHA Stroke logit scores for each person post-stroke.

4.2.7 Data Analyses

Algorithm outputs for the similarity value were examined using a two-way analysis of variance (ANOVA) to assess performance in persons post-stroke vs. controls across the three task types (U, BS, BA). This two-way ANOVA was run separately for each of the 10 sensor-to-sensor comparisons based on sensor location. We present results from three (of ten) primary sensor-to-sensor comparisons that represent the most frequently observed compensatory movement patterns during upper limb tasks [138, 245, 294] and as such were expected to be differentially sensitive not only to gesture type, but to the types of impairments present in individuals post–stroke. These key sensor comparisons shown in **Figure 4.5**, were: sternum to wrist 1 (S-W1), wrist 1 to wrist 2 (W1-W2), and wrist 1 to upper arm 1 (W1-UA1) for all three task types for each group. Thus, potential compensatory patterns assessed based on sensor comparison included trunk involvement during arm motion (S-W1); limb disuse or asymmetry during bimanual tasks (W1-W2), and movement efficiency or atypical synergistic motion during any task type (W1-UA1). A comparison of similarity values between task type for the three key sensor pairs among each group were further examined using a Tukey-Kramer post-hoc analysis.

The arm use ratio was compared to scores from the UE-FMA and Ad-AHA Stroke from persons post-stroke, using a predictive linear regression analysis to determine the extent to which the use ratio of raw wrist sensor data extracted during task performance could predict scores on each clinical assessment.

4.3 Results

Forty participants were recruited; 20 individuals post stroke (53.5 $\pm$ 11.5 yrs) and 20 age-matched controls who had not had a stroke (52.3 $\pm$ 12.5 yrs). Individual demographics and clinical features of participants with hemiparesis post-stroke are shown in Table 1 with means (SDs) listed for controls. All participants were included in the analyses, except for the comparison of the arm use ratio against the UE-FMA and Ad-AHA Stroke which only included findings from participants post-stroke.

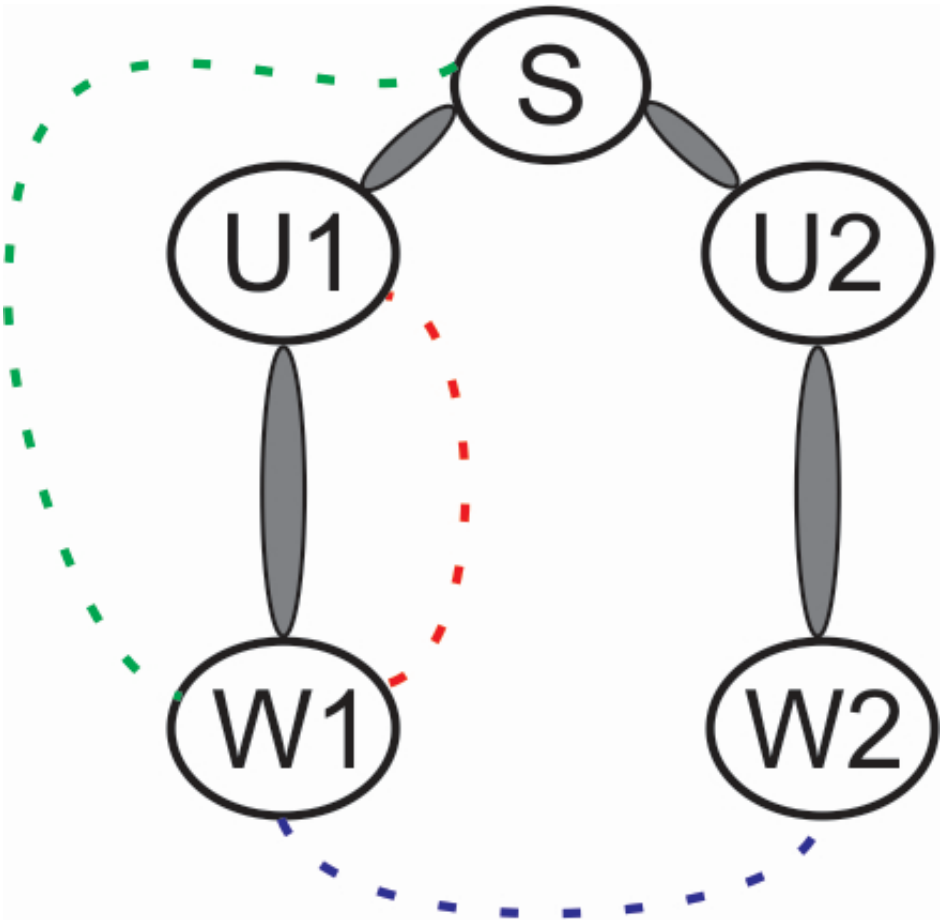


Figure 4.5: Three key sensor to sensor comparisons: S-W1 (green line), W1-W2 (blue line), and W1-U1 (red line).

Differences in similarity value between groups (controls and persons post-stroke) across task type (U, BS, and BA) were found. **Figure 4.6** depicts the similarity values for one representative person post-stroke and one control for the three key sensor comparisons: S-W1, W1-W2, and W1-U1 across the three types of tasks. The variability in similarity between sensor pairs are represented by the colormap (dark red = high sensor similarity and limb coupling; dark blue = low sensor similarity and limb coupling). The relationships for group across three task types based on the similarity score for three sensor comparisons are shown in **Figure 4.7**. As expected, we found group and task inter- and intralimb differences. Main effects for group were found for two sensor comparisons: S-W1, ($F=55.33$, $p<0.01$); and W1-W2 ($F=143.28$, $p<0.01$). Main effects were also found for task type for two sensor comparisons: S-W1 ($F=20.17$, $p<0.01$) and W1-W2 ($F=183.23$, $p<0.01$). A significant group by task interaction was found for all three sensor comparisons: S-W1 ($F=9.73$, $p<0.01$); W1-W2 ($F=90.54$, $p<0.01$); and W1-U1 ($F=6.22$, $p<0.01$).

Post-hoc analyses of the similarity values for controls for the three key sensor pairs against the three task comparisons revealed significant differences for 8 out of 9 instances (**Fig. 4.8**). For the U-BS and BS-BA task comparisons: S-W1 and W1-W2 significantly differed ($p<0.01$ respectively); and W1-UA2 significantly differed ($p<0.01$). For the U-BA tasks, the similarity values among two sensor pairs differed significantly: S-W1 ($p<0.01$) and W1-W2 ($p<0.01$). For persons post-stroke, only 3 out of 9 instances were significantly different. For U-BS tasks: S-W1 differed ($p<0.01$); and W1-UA2 differed ($p<0.01$), whereas, for BS-BA tasks: only S-W1 differed ($p<0.01$).

We examined the relationship between the use ratio (paretic/non-paretic arm) for all three modalities and scores on the UE-FMA and AdAHA Stroke. As shown in **Figure 4.9A** the use ratio was significantly predictive of scores on the UE-FMA based on the integral for acceleration ($R2=0.67$, $p<0.01$), angular rate of change ($R2=0.67$, $p<0.01$), and orientation ($R2=0.46$, $p<0.01$). The use ratio was also predictive of scores on the Ad-AHA Stroke (**Fig. 4.9B**) based on the integral for acceleration ($R2=0.60$, $p<0.01$); angular rate of change ($R2=0.55$, $p<0.01$) and orientation ($R2=0.52$, $p<0.01$).

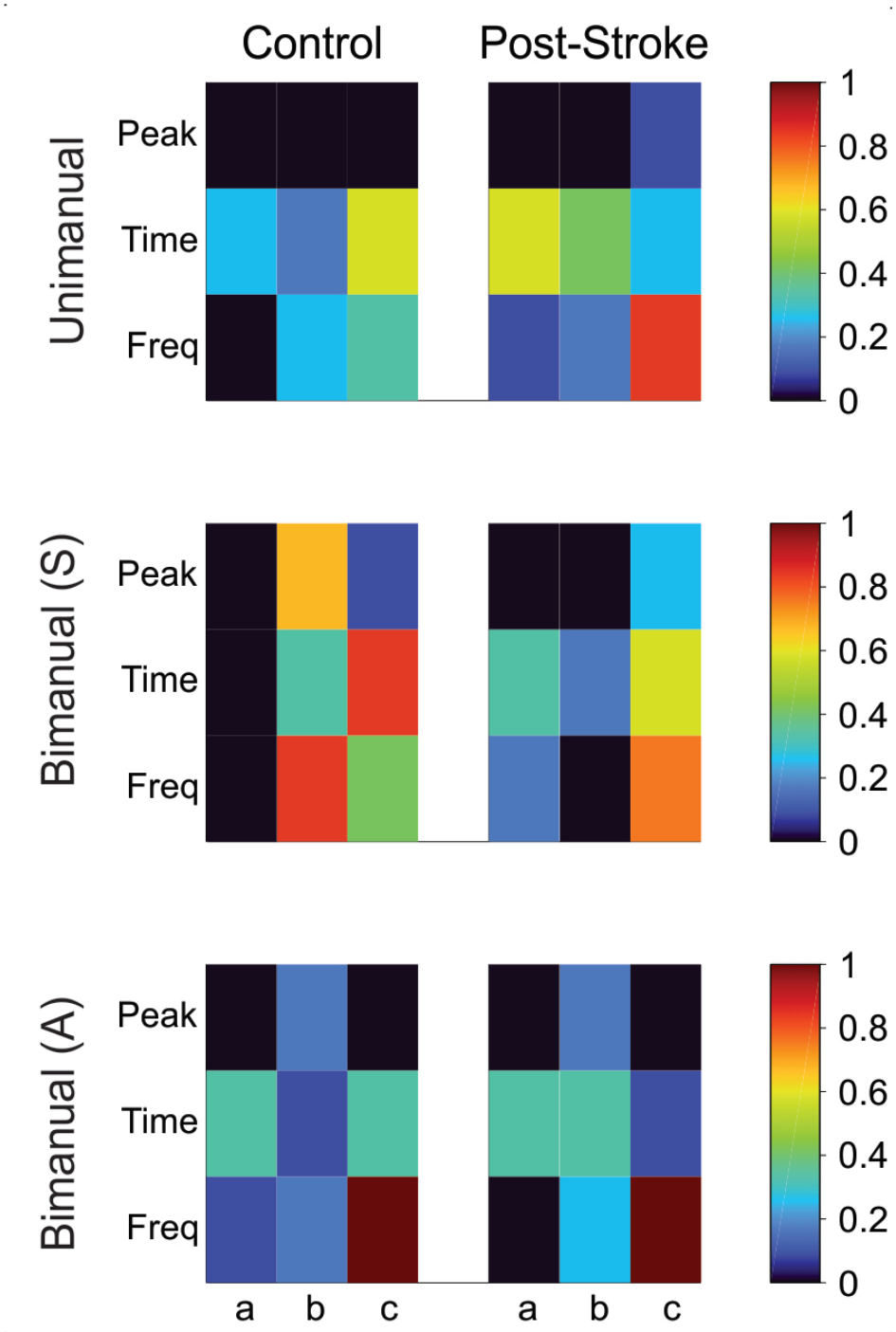


Figure 4.6: Similarity metrics for representative participants (Post-stroke – 19, 46 years old, UE-FMA-30; and Control 5, 49 years old) for the 10 sensor-to-sensor comparisons by task type: Unimanual (U), Bimanual Asymmetric (A), and Bimanual Symmetric (S). The comparisons were: a) sternum to wrist 1 (S-W1), b) wrist 1 to wrist 2 (W1-W2); and c) wrist 1 to upper arm 1 (W1-U1). Dark red color blocks indicate similarity (correlation) was closer to 1.0; dark blue color blocks indicate similarity closer to 0.

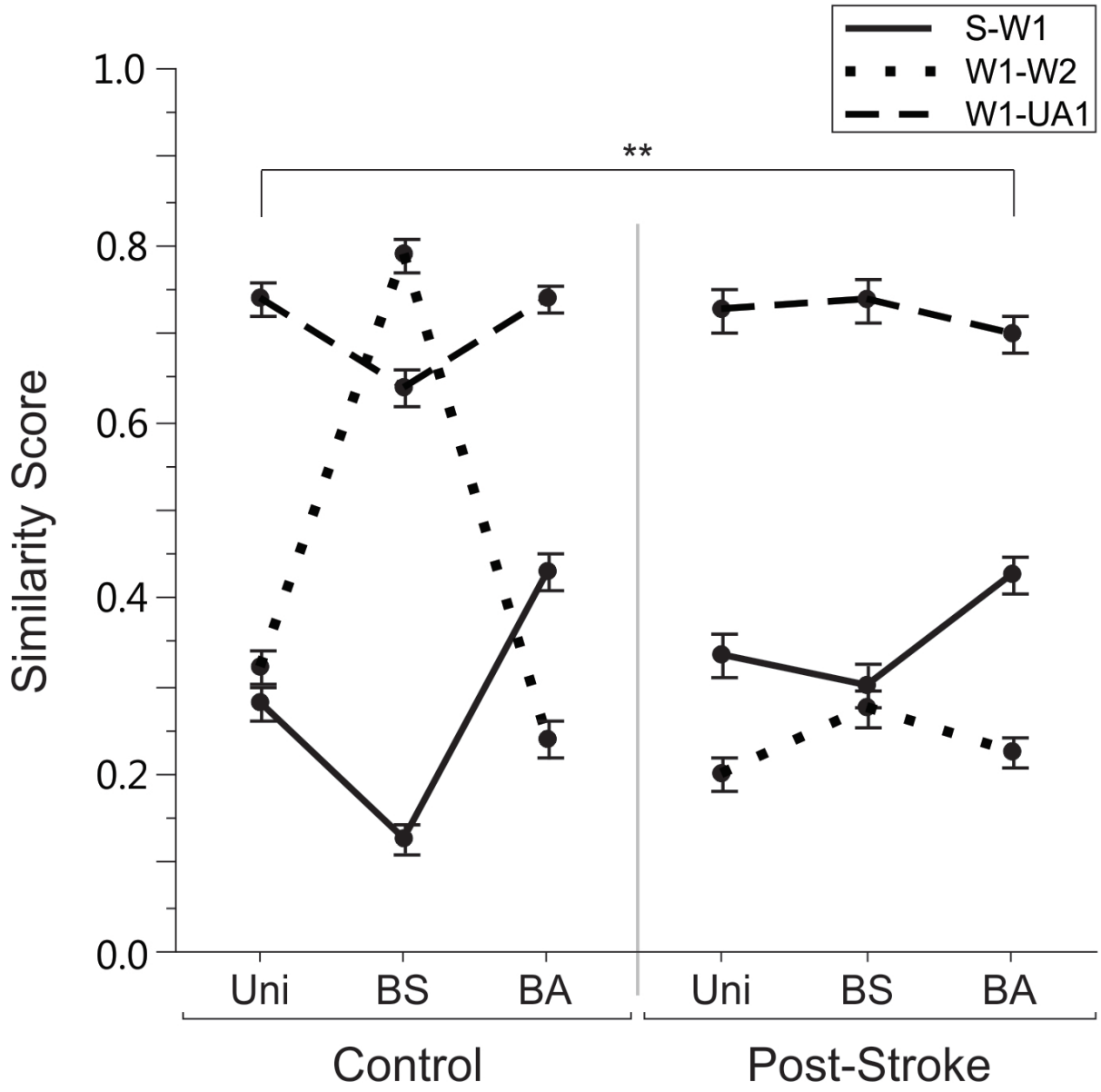


Figure 4.7: Mean (SEM) similarity metrics by group for each task type (unimanual , bimanual symmetric, and bimanual asymmetric) and the three sensor-to-sensor comparisons: Wrist to sternum (S-W1), wrist 1 to wrist 2 (W1-W1), and wrist 1 to upper arm 1 (W1-U1).

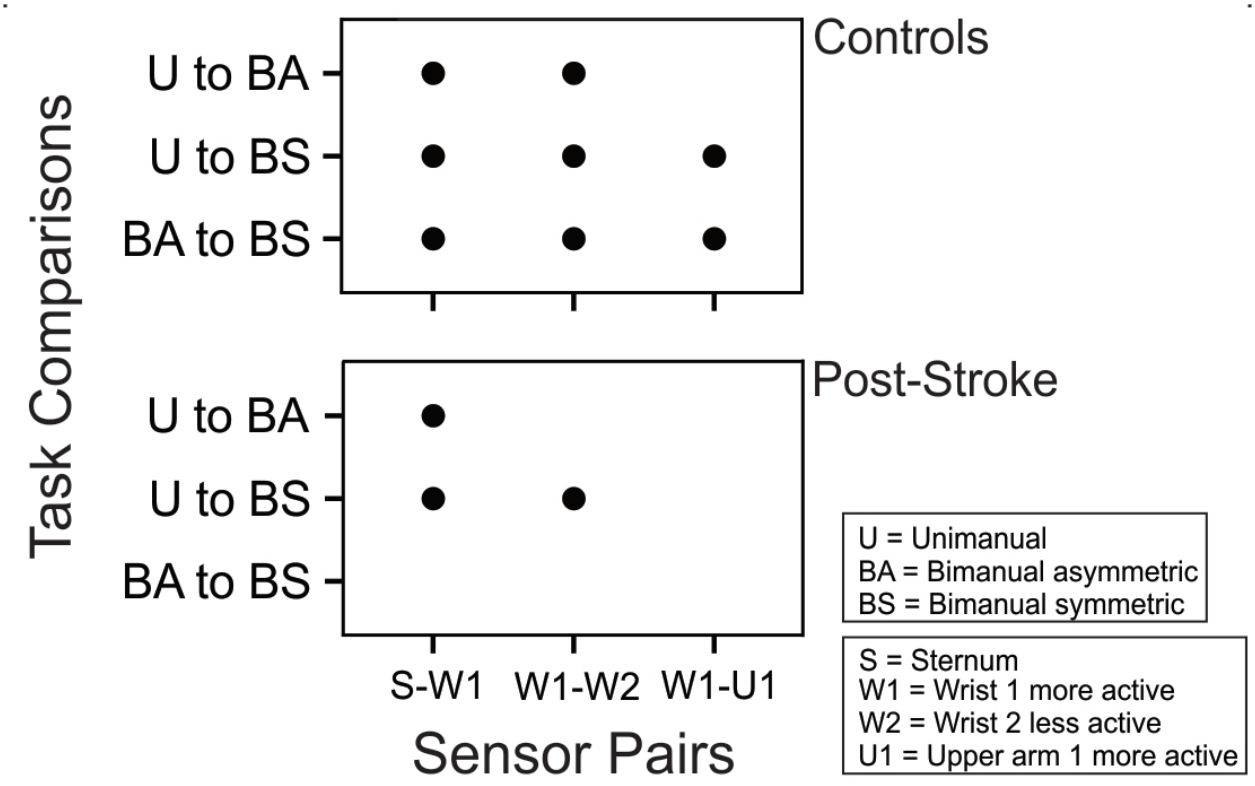


Figure 4.8: Significant differences from the within group, post-hoc analyses between the three different sensor locations and three different task types for healthy controls and persons post-stroke based on sensor comparisons. Red dots for controls and blue dots for persons post-stroke represent significant differences at $p<0.01$.

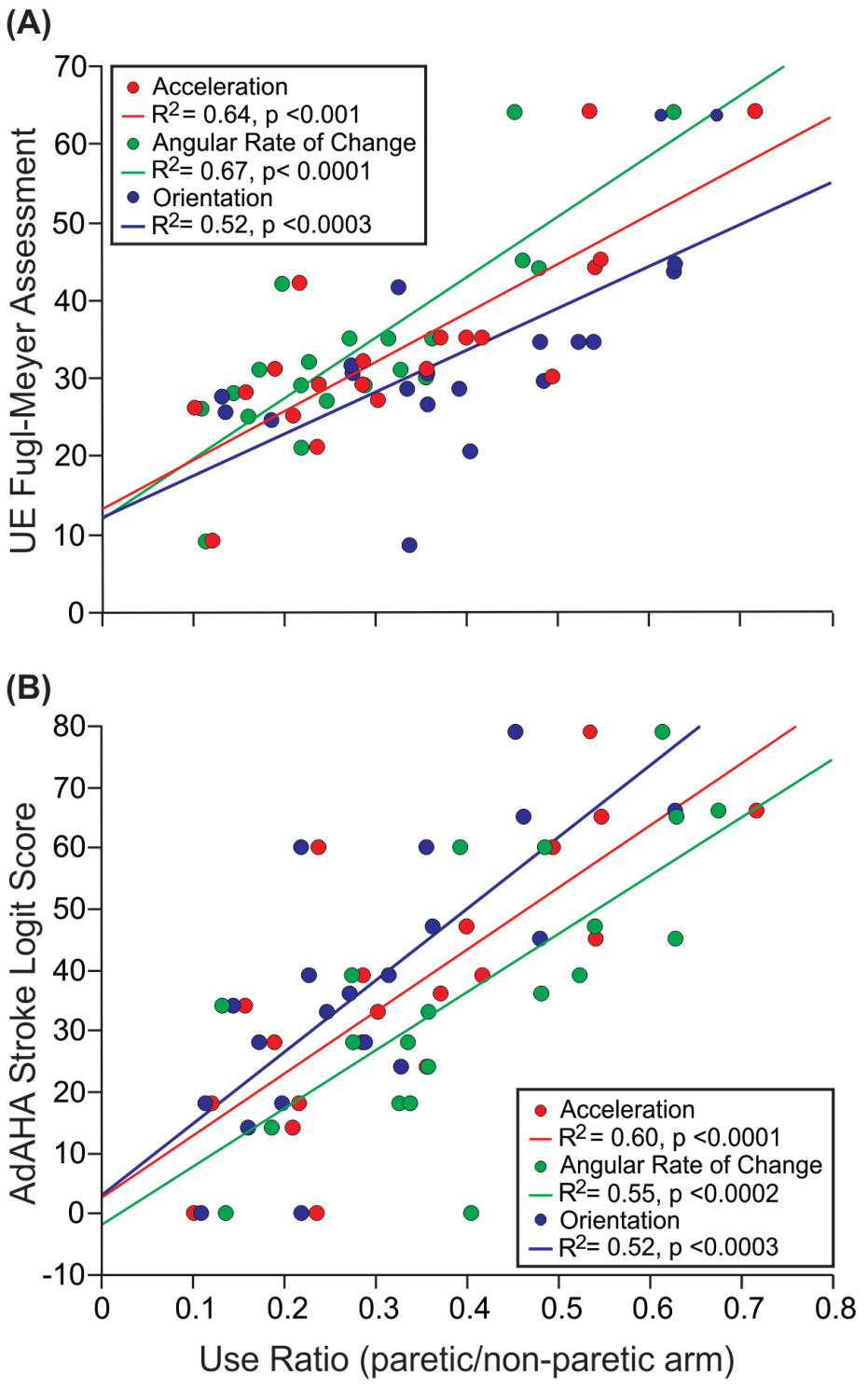


Figure 4.9: Arm Use Ratio based on the integral from wrist sensor data (paretic/non-paretic arm) for acceleration, angular rate of change, and orientation associated with A) UE Fugl Meyer Score; and B) Ad-AHA Stroke logit score.

4.4 Discussion

In this chapter, we sought to quantify intra- and interlimb coordination during performance of unimanual and bimanual tasks through the use of a novel algorithm comparing performance between individuals post-stroke and controls. The findings suggest that our sensor-based algorithm accurately discriminated between U, BS, and BA tasks in persons post-stroke and controls. It also differentiated performance between the two groups within tasks. Limited differences in persons post-stroke for the key sensor-to-sensor comparisons between task type suggests that there was greater intralimb and interlimb coupling than that of controls for most tasks. Furthermore, the use ratio was found predictive of scores on the UE-FMA and Ad-AHA Stroke.

Spatial and temporal coordination between the limbs is essential for bimanual task performance. During symmetric tasks, one limb typically mimics the other with regard to speed and movement pattern causing a mirror-like effect. Similar interlimb performance for symmetric tasks has been reported in healthy adults during bilateral pointing and lifting tasks [114]. In children and adults with hemiparesis, slower movement of the paretic limb typically slows performance of the less affected limb resulting in similar movement duration with compensatory kinematics [63, 262]. Here, we were able to detect typical performance as well as compensatory strategies by examining the similarity score between sensors. During symmetric tasks, such as folding a towel or donning a hat, there were much higher similarity scores for the two wrist sensors among controls than for persons with post-stroke hemiparesis. In persons with hemiparesis there were higher similarity scores between the more active wrist and the sternum than for controls during tasks classified as bimanual symmetric. One likely kinematic explanation for this result is the incorporation of both trunk and arm movement during task performance, a strategy often used in individuals post stroke to compensate for incomplete or weak shoulder flexion or elbow extension [245, 50].

During asymmetric tasks, the role of each limb is typically differentiated yet performance of one limb may be affected by the constraints faced by the contralateral limb [165]. Many studies have used the drawer-opening task [113, 278] to examine asymmetric task performance. In healthy children and adults, the act of opening the drawer with one hand and

reaching to pick up a peg with the other occurs almost simultaneously regardless of the role of each hand [63, 97]. Yet, in children and adults with hemiparesis, the actions between limbs for this task are typically more sequential [63, 97] and performance differs depending on the role of each limb. In this study, the asymmetrical tasks typically involved object stabilization with one hand and movement of an object with the other hand. Interestingly, for these BA tasks, the similarity scores for the three key sensor comparisons did not significantly differ between groups. Thus, the combined metrics gathered from the motor behaviors of reaching, holding, and moving for the bimanual asymmetric tasks were generally similar between limbs among both groups.

Studies examining UE intra- and interlimb coordination have often used wrist-based sensors to record kinematics during arm movement [63, 262, 113, 227]. The use of two sensors allows for the calculation of an arm use ratio allowing the examination of changes in arm function during recovery or over the rehabilitation process. The relationship between the use ratio (based on the integral of sensor modalities) and the UE-FMA found in this study is consistent with existing literature [63, 258, 59], suggesting that the tasks we used are representative of functional tasks known to be sensitive to post-stroke impairment and function. Unique to this study is the finding that the arm use ratio was also predictive of Ad-AHA Stroke logit scores. Given the sensitivity of the UE-FMA motor domain and Ad-AHA Stroke to limb coordination, this relationship provides preliminary support for the validity of this sensor-based metric as a method to examine intra- and interlimb coordination.

4.4.1 Limitations

The amount and type of arm use in persons with hemiparesis often varies for similar bimanual tasks. Additionally, although results for performance based on the similarity values were consistent within task, our study had participants perform two repetitions each of six different tasks. Future work will incorporate a greater number of tasks, and will include longitudinal assessment to assess responsiveness of the measures. As Figure 3 reveals, although between group differences were statistically significant, this was not necessarily the case for each task type among the three sensor comparisons, particularly for persons post-stroke whose similarity values were similar among task type for each sensor comparison. Additional

studies, with larger samples and more tasks, may reveal greater task differences. Finally, participant behaviors in monitored (e.g., lab and clinical) settings may vary from those performed in ambient settings. This has been shown for a variety of behaviors and in clinical and non-clinical populations [50]. Full implementation of our algorithm to monitor changes in intra- and interlimb coordination will require separate studies of the variability of our features when extracted in ambient settings.

4.4.2 Implications

The use of five sensors during task performance allowed us to analyze intralimb and interlimb coordination. This level of analysis can be advantageous when quantifying change in coordination of the more affected limb during unimanual and bimanual tasks. Our quantitative assessment revealed the use of compensatory mechanisms during unimanual and bimanual task performance, such as incorporation of trunk movement during reaching and a reduction in acceleration and movement amplitude. Thus, our algorithm-based assessment has the potential to quantitatively assess coordination within and between limbs in persons with hemiparesis during task performance, to allow monitoring of capability and change with recovery or following rehabilitation. Metrics of intra- and interlimb coordination provided here provide a quantitative measure of performance during everyday tasks, which may be of value in future clinical trials.

In **Chapter 3**, we revealed that compensation is often described in the context of a task being perform. The novel algorithm we developed in this chapter was capable of quantifying interlimb coordination during performance of unimanual and bimanual tasks. We also validated our findings by showing that sensor derived data correlates with common clinical measures. Having shown that features derived from IMU data are sensitive to task type, in the following chapters, we will explore this data to begin autonomously classifying tasks using machine learning methods. Autonomously classifying tasks is key to recognizing and reversing compensation in real-time in the ambient setting.

Chapter 5

Comparison of Machine Learning Approaches for Classifying Upper Extremity Tasks in Individuals Post-Stroke

5.1 Introduction

Machine learning is broadly defined as "computational methods using experience to improve performance or to make accurate predictions [182]." Machine learning applications are particularly useful in the field of medicine where, because of the complexity, it can be difficult to "capture the relevant information in rules [220]" since machine learning models learn from examples instead of being programmed with rules [207]. Machine learning applications can be applied to every step of the medical process from prognosis [24], to diagnosis [186], to treatment [70, 219, 81], and they can be used to improve clinician availability [45, 79, 233] and workflow [134].

In recent years, machine learning methods have been applied specifically to stroke in: detection [266, 159], diagnosis [209], treatment [154], outcome prediction [31, 90], and prognosis evaluation [141, 9, 10]. In a review of artificial intelligence (AI) in healthcare,

Jiang et al. [154] state that AI will help with answering more complicated and authentic clinical questions which will lead to better decision making in stroke management, and [105] indicate that researchers have already begun to make promising strides in this direction [295].

There are seven common types of machine learning methods, but in this analysis, we are focused on three that are commonly used in biomedical and post-stroke applications: supervised, unsupervised and reinforcement learning [182]. For supervised learning, an algorithm is trained using labelled datasets to classify data or predict outcomes. As the model receives data, its weights are iteratively adjusted in a cross validation process until the model fits the data [26]. Supervised learning is often used in classification, regression and ranking. Some common examples of supervised learning methods are decision trees, random forests, k-nearest neighbors (KNN), and logistic regression. Unsupervised learning requires only input feature values and the algorithm discovers hidden structure in the training data based on the given features [26]. Unsupervised learning is often used in clustering and dimensionality reduction, and some common examples of unsupervised learning include k-means and hierarchical clustering. Finally, reinforcement learning is training a machine learning model (or agent) to learn from its own experiences using trial and error [228, 68]. This is different than supervised machine learning, where feedback consists of the "correct" set of actions to perform or complete a task, in that reinforcement learning uses "rewards" and "punishments" for positive and negative behavior respectively. Unlike unsupervised learning, where the goal is to find similarities between data points, the goal of reinforcement learning is to maximize the reward of the agent. An example of reinforcement learning is a Markov decision processes. While other machine learning methods are available, the three described are usually sufficient to cover most common healthcare applications (such as description, prediction, and casual inference) [106, 68].

In **Chapter 3**, we discussed how compensation is defined in the context of both a segment used and a task being performed when compensation is identified. We also demonstrated, in **Chapter 4**, that there is a significant difference between post-stroke and control groups as well as between task types using features derived from body worn inertial measurement units (IMUs). In this chapter, we will use the features and practices from **Chapter 4**, as well as other selected features, to explore several machine learning methods used for task

classification for people post-stroke and age-matched controls. Several supervised machine learning methods will be used to attempt to classify different windows and features derived from the data. While unsupervised learning was used to cluster the data and perform dimensionality reduction for other learning methods, it was not included in the final analysis. We used unsupervised learning and other methods to investigate underlying trend in the data that the extracted features were not sensitive to, but the methods evaluated did not increase classification accuracy beyond supervised learning. This could be because of the varied nature of the human subjects data collected, or it could be that the data are heavily weighted toward certain tasks and the unbalanced nature of the data makes accurate classification difficult.

5.2 Data Collection and Processing

5.2.1 Participants

Twenty persons post-stroke and twenty age-matched controls were recruited from a convenience sample. The study population consist of the participants mentioned in **Chapter 4**. Participants were recruited according to the rules of the relevant Institutional Review Boards of Chapman University, Columbia University Medical Center, and Teachers College, Columbia University. Inclusion criteria for participants post-stroke were hemiparetic stroke causing weakness primarily on one side of the body, and in the chronic stage of recovery (>6 months post-insult).

5.2.2 Design

Participants were asked to perform a series of tasks which belonged to one of five categories. These categories included: UL, unimanual tasks performed with the dominant (control) or less-affected (post-stroke) hand (e.g., reaching for a bottle or spoon); UM, unimanual tasks performed with the non-dominant or more-affected hand; BAL, bimanual asymmetric tasks actively performed with the dominant or less-affected hand (e.g., opening a bottle or stirring with a spoon in a bowl held by the contralateral hand); BAM, bimanual asymmetric

tasks actively performed with the non-dominant or more-affected limb; and BS, bimanual symmetric tasks where both limbs contribute equally in active mirrored movements (e.g., folding a towel or donning a hat). For a full list of tasks, see [178].

5.2.3 Procedure

Prior to task performance, participants were outfitted with five APDM Opal inertial measurement units (IMUs) which measured triaxial acceleration, angular rate of change, and magnetic field strength at 128 Hz [5, 275]. These sensors were secured via straps or a harness to the left and right wrists, the left and right upper arms (just above the elbow), and the sternum. Participants were filmed during task performance to determine start and stop times for each task. Start and stop times were then extracted using video editing software.

Sensor data were recorded and saved onto a single compressed HDF5 file for each participant, which was then extracted in MATLAB and transferred to a structure for analysis. Initially, all data were smoothed using a band-pass Butterworth filter with a frequency band of 0.5 to 2.0Hz. These frequency cutoffs were selected as they are appropriate for signal power of volitional movement typical of ADLs [188]. Data were then detrended to remove any unwanted effects from gravity, drift, or obtrusive magnetic fields. Following data pre–processing, individual tasks were segmented using the video-based start and stop times. This procedure resulted in a data set consisting of twelve total tasks (2 UL, 2 UM, 2 BAL, 2 BAM, 4 BS), 40 participants, and 45 signals (five sensors, three modalities, and three axes).

5.3 Supervised Learning

During supervised learning, a model is trained using labelled datasets, which can can consist of raw data or derived features, to classify data. When using continuous time series data, it is common to extract representative features from sections, or windows, of data to increase computational efficiency [116]. In this section, we will perform feature extraction on windows of data to create the datasets for supervised learning. To decrease computational load and time, feature selection will be performed to remove any correlated or overlapping features. Window size, features selected, and number of features will be iterated, and the resulting

datasets will be evaluated using a variety supervised learning methods to classify the data. Results of this analysis will be shown in **Section 5.3.2** below.

5.3.1 Methods

Comprehensive Feature Extraction

The set of features extracted from the data set were selected based on [116] and [69] as they indicate features which are derived from time-series data collected by body-worn IMUs and useful for classifying human motion. Our feature set included: mean, median, variance, standard deviation, minimum, maximum, range, root-mean-square, correlation, cross-correlation, tilt angle, zero-crossings, signal magnitude area (or average), signal magnitude vector, wavelet transform (sum of coefficients). In addition, Fourier transform (AC component), Fourier transform (DC component), Fourier transform (dominant frequency), spectral energy, spectral entropy, Euclidean distance, and dynamic time warping were included. As mentioned, having too many features can increase computation load and time required. Using too many features can also result in having features which are correlated and can lead to over-fitting [276, 274]. Given the number of features and the data set size, we then employed a feature selection approach.

Feature Selection

Selection was done using the Weka package, a suite of machine learning tools [83]. Weka provided a number of different algorithms for feature selection (or attribute evaluators), which could employ any of three different search methods to determine features most useful in classification or regression. We chose attribute evaluators which employed the Ranker search methods, which ranked attributes by their individual evaluations. These attribute evaluators included: 1) Classifier Attribute Evaluator which "evaluated the worth of an attribute by using a user-specified classifier" (default being ZeroR classifier), 2) ReliefF Attribute Evaluator which "evaluated the worth of an attribute by repeatedly sampling an instance and considering the value of the given attribute for the nearest instance of the same and different class," 3) Info Gain Attribute Evaluator which "evaluated the worth of an

attribute by measuring the information gain with respect to the class," and finally 4) Gain Ratio Attribute Evaluator which "evaluated the worth of an attribute by measuring the gain ratio with respect to the class." The four methods for attribute evaluation provided results that varied slightly from each other, so a conglomerate of the four methods was created to provide one comprehensive, ranked list of all extracted features. ASCII-based data files in the attribute-relation file format (.arff) were created using the top 10, 20, 30, 40, 50, 100, 150, 200, 300, 400, 500, and 660 (the maximum number) features from the ranked list, where each column represented an individual feature and each row represented an individual participant and task.

Classification

Using the various feature sets, the next step was to evaluate classifier performance. Features were separated such that the two tasks (or four in the case of the bimanual symmetric tasks) and 20 participants were treated as repeated measures (the two groups were analyzed separately). This meant that, for four of the five task types, there were a select number of features repeated forty times and for the fifth task type, bimanual symmetric, the select features were repeated eighty times. Each data file was pre-processed in Weka and evaluated using 9 different classifiers with a random, 10-fold cross-validation. Cross-validation was used to determine if the test criteria can be generalized across samples [202]. These classifiers, chosen based on [29], were: Naïve Bayes, SMO, IBk, KStar, Multiclass Classifier, Bagging, Decision Tree, J48 and Random Forest. Tuning classifiers is a separate research domain and therefore, not discussed in this paper.

Static Data Windows

When processing continuous data, the segment of data selected for analysis, or *window* of data, has relevance for computational cost and processing speed (see [29]). The size of these windows affects the classification performance [184]. For our data, we chose a variety of windows to span each gesture. This approach resulted in the generation of more data and increased sample size. From an application perspective, real-time analyses (the eventual utility of the proposed approach) must analyze and characterize continuous data. These data

do not have clear start and stop times, so any classifier must either automatically detect start and stop times or classify a task type based on a static window length, regardless of the true amount of time required to complete the task. Ultimately, we divided the data into 1 and 2 second windows for analysis. These window lengths were chosen because the shortest full gesture for both post-stroke and control participants was approximately one second and because shorter window lengths would be too computationally expensive to attempt based on the number of features and repeated measures. Once the data were windowed, they were randomly divided (in Weka) using a 10-fold cross-validation strategy to test the classifiers.

One potential problem with this approach stemmed from the random separation of participant data. For example, a task that took four seconds to complete might have been broken into two two-second windows, and one window might have been used to train the model while the other window used to test the model. Thus, the generation of a participant-specific model may have possibly presented satisfactory results for participants on which the model was trained, but not for others.

Hand-Picked Feature Extraction

The extracted features discussed so far were derived from [83]. While all these features can be useful in classifying human movement, a much smaller list of metrics are commonly used in ADL classification. To evaluate classifier performance and compare it to the approach used in [29], we separately hand-picked 8 features from each sensor, axis, and modality: Root Mean Square (RMS), Mean, Signal Magnitude Area (SMA), Signal Vector Magnitude (SVM), Energy, Entropy, FFTPeak, and Standard Deviation (STD). Using these eight features, we ran the attribute selection again, this time having divided into data sets with the top 10, 20, 30, 40, 50, 100, 150, and 210 features (210 being the maximum after features from each sensor, axis, and modality were extracted). We then reran the nine different classification methods having used a random 10-fold cross-validation strategy for both groups to determine classifier performance for the hand-picked features.

Overlapping Data Windows

In addition to data windowing, we also included data over-lap, also known as a window shift. Overlapping windows was a successful approach to increase the number of correctly classified instances in gesture recognition and classification [29]. We investigated windows with overlaps of 25%, 50%, 75%, and 90%. Due to the increased number of windows, analyses were performed on data sets with 50 or fewer features. Again, data were separated using random, 10-fold cross-validation.

To evaluate participant-independent models, we then separated the data by participant and tested the process by randomly removing two of the twenty participants and training on the remaining eighteen. The two that were left out were then used to test the model to determine the accuracy based on correctly classified instances. This process was repeated ten times to achieve a 10-fold cross-validation.

5.3.2 Results

Selected Features

Our first objective was to determine whether all 660 features extracted from the data were necessary for accurate classification. The number of features used to build each classifier was incrementally increased as described above, and we selected the appropriate number of features from the top of the ranked list that was generated using Weka's attribute ranker. The selected features were then used to test each classifier, and we performed a 10-fold cross validation to evaluate the number of correctly classified instances. The results of this cross validation for participants post-stroke and controls are shown in Figure 5.1.

For persons post-stroke, the peak correctly classified instances (59.17%) resulted when the classifier was built using the top 200 selected features. For the control group, the maximum correctly classified instances (79.17%) came from the Random Forest classifier while the top 40 features were used to build the classifier. The number of correctly classified instances (**CCIs**) trended toward a peak value after the number of features used to build the classifier reached 50 or 100, and then the number of CCIs trended to plateau as the number of features used increased past that peak.

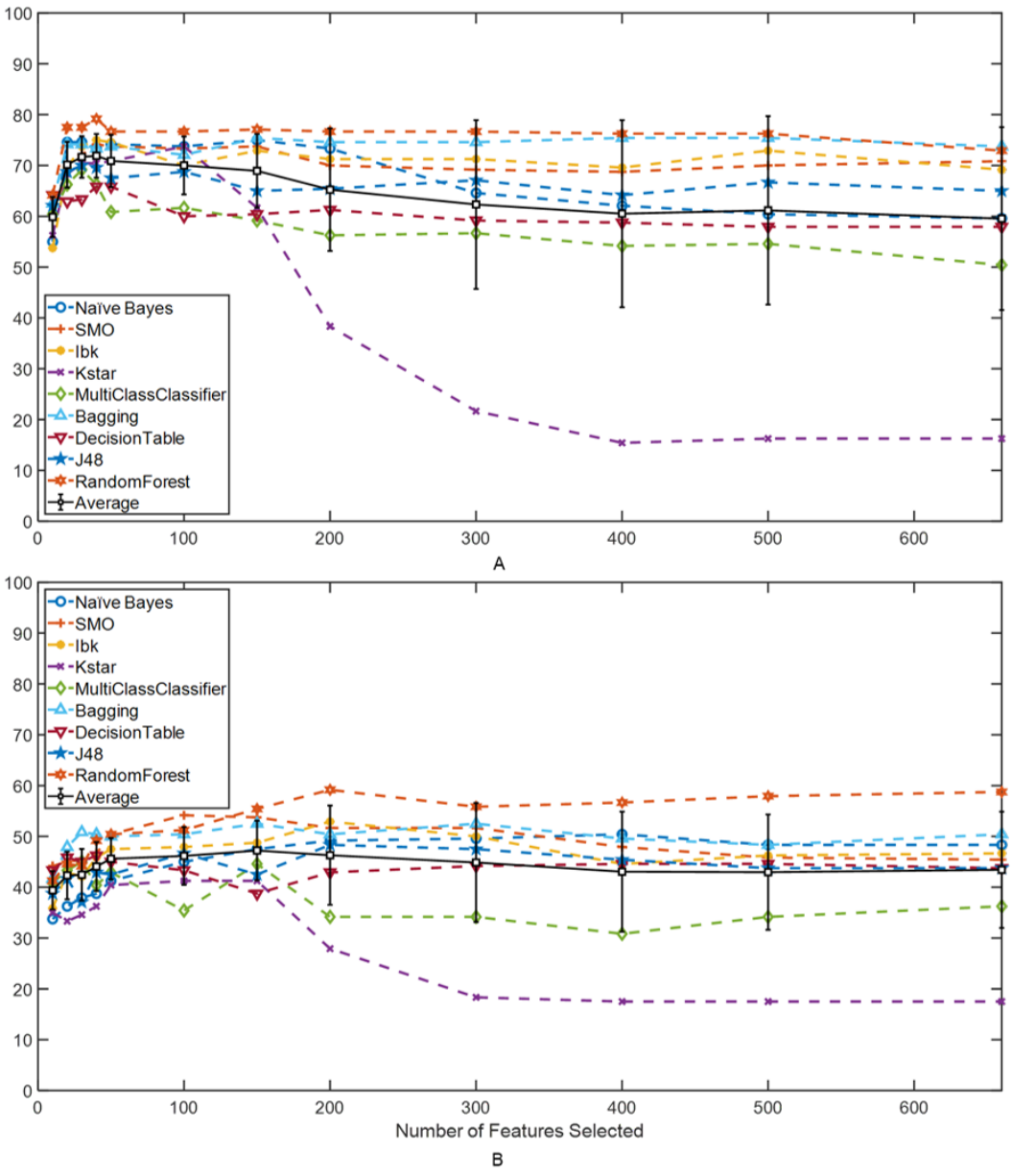


Figure 5.1: Correctly Classified Instances versus Number of Features Selected for Control (A) and Post-Stroke (B) Populations. The results are shown for the nine classifiers used, as well as an average of all nine classifiers.

Windowing Data

The effects of windowing are depicted in Figure 5.2. For the post-stroke group, there was little change in the number of correctly classified instances, regardless of the change of window size. For the control group, the number of CCIs decreased significantly as the window size decreased to two seconds, then increased slightly when the window size increased to one second. The trend for the number of CCIs plateauing as the number of features selected increases persisted, even as the window size changed.

Limited Feature Selection

The effect of using the hand-picked features when compared to using the comprehensive feature set is depicted in Figure 5.3. It shows the number of correctly classified instances when models are built using the top 10, 20, 30, 40, and 50 features of the full feature set compared to the limited feature set. These features were drawn from 2 second windows. For both groups, the change in the number of correctly classified instances was not statistically significant.

Overlap of Windowed Data

Figure (Fig. 5.4) shows the number of correctly classified instances compared to the overlap of the two second windows that were used to segment the data. The models were built using the top 50 features selected from the hand-picked features. The percent overlaps tested were 0, 25, 50, 75, and 90. As the percent overlap increased, the number of correctly classified instances increased for a majority of the classifiers in both groups. For the persons post-stroke, at 90 percent overlap, the IBk classifier achieved 95.6% CCI and the Random Forest achieved 94.4% CCI. For controls, the IBk classifier achieved 91.4% and the Random Forest achieved 89.9% CCI.

User-Independent Model

The final analysis was to determine how many of the CCIs were influenced by oversampling in user-dependent models. Weka used a random 10-fold cross-validation, so with a 90%

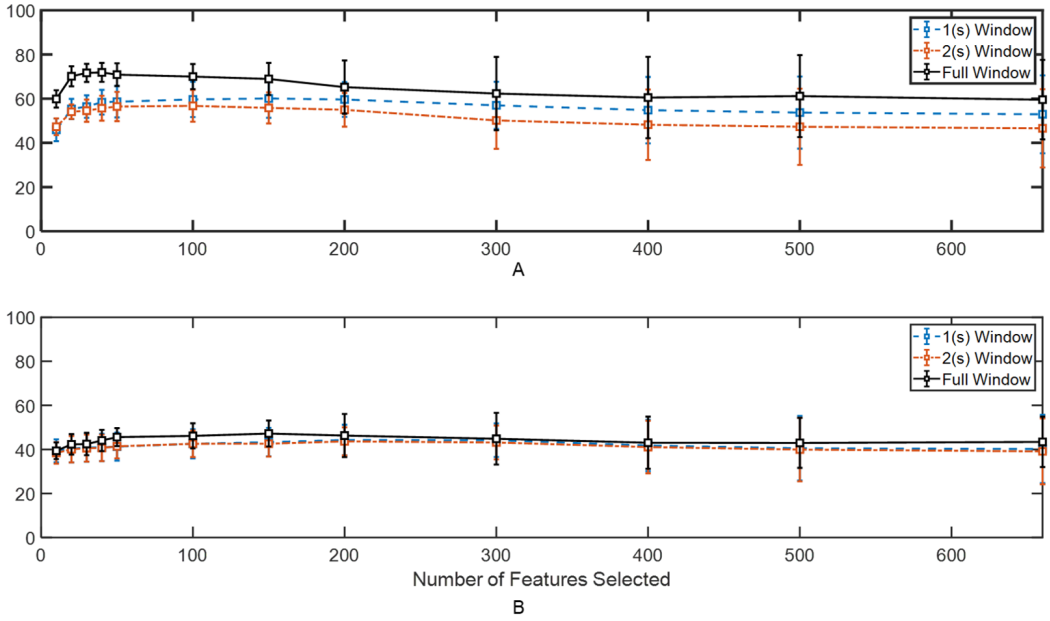


Figure 5.2: Correctly Classified Instances versus Number of Features Selected for Control (A) and Post-Stroke (B) Populations. The average of the nine classifiers is shown for one and two second windows, as well as the full task length.

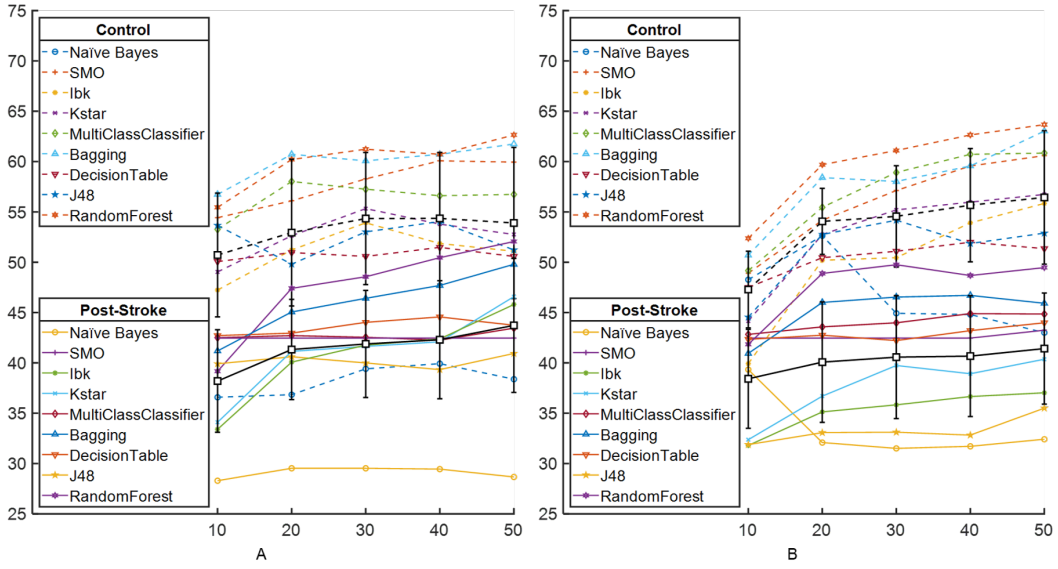


Figure 5.3: Correctly classified instances when hand-picked features were selected (A) compared to selecting features from the full extraction process (B). The nine classifiers and their averages are shown for both Post-Stroke and Control groups for each feature selection process.

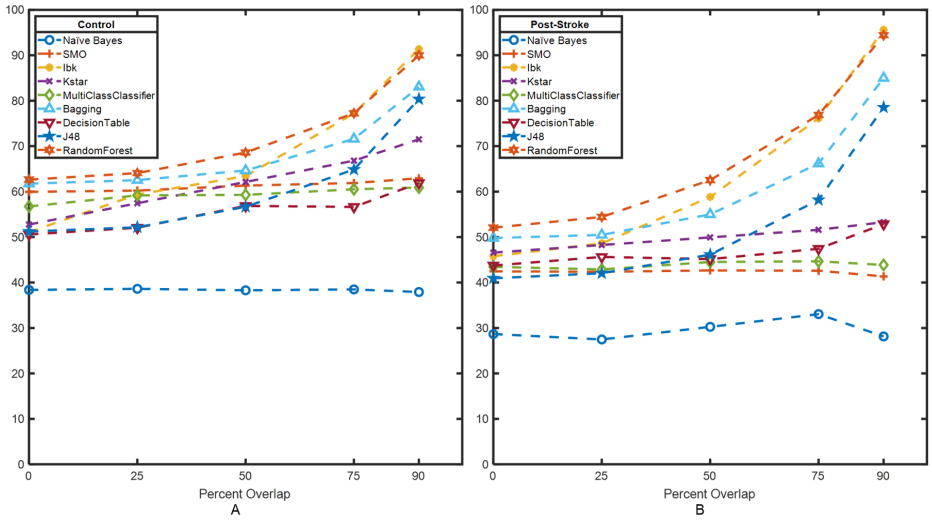


Figure 5.4: Correctly classified instances as percent overlap of the two second window increased for Control (A) and Post-Stroke (B) groups.

overlap, it was likely that the model was being trained on data that was similar to the data that was being used to test the model. The results are shown in Figure 5.5. For participants post-stroke, at 90 percent overlap, the IBk classifier decreased from 95.6% CCIs to 31.5% CCIs and the Random Forest decreased from 94.4% CCIs to 38.3% CCIs. For controls, IBk classifier decreased from 91.4% to 55.4% and the Random Forest decreased from 89.9% CCIs to 61.8% CCIs.

Discussion

Our goal was to investigate the application of machine learning methods to minimally intrusive sensor data in order to classify categories of ADLs. Given the between subject variability within task types and the limited sample size, it is reasonable to expect that there would be difficulty in generating a generalized model that is capable of differentiating between the five task types for any participant. However, we demonstrated that, given training data from a participant, a user dependent model can be generated which produces satisfactory results.

With respect to feature extraction, our results demonstrated that classification accuracy increases with the number of features up to a point. After reaching 50 selected features, the classification accuracy plateaus. Although classification accuracy only reached approximately 80% for controls and 60% for participants post-stroke, these results indicate that future models do not need a large number of features to achieve peak accuracy. This is of particular importance for real time application as the fewer features extracted and used, the less processing and computational load required.

In examining the potential of classifying tasks using data extracted from a static window of data in real time, we found a significant decrease in classification accuracy when compared to extracting features from the full task window. This indicates that over the course of a task performance, there are features not captured or recognizable in a windowed segment of the data, or at least are not sufficiently distinct, given the variability between participants, for the models to be capable of distinguishing between tasks. This suggests a need for an automated process for determining the beginning and end of task performance, so a task in full can be extracted and features from an entire task can be analyzed in real time.

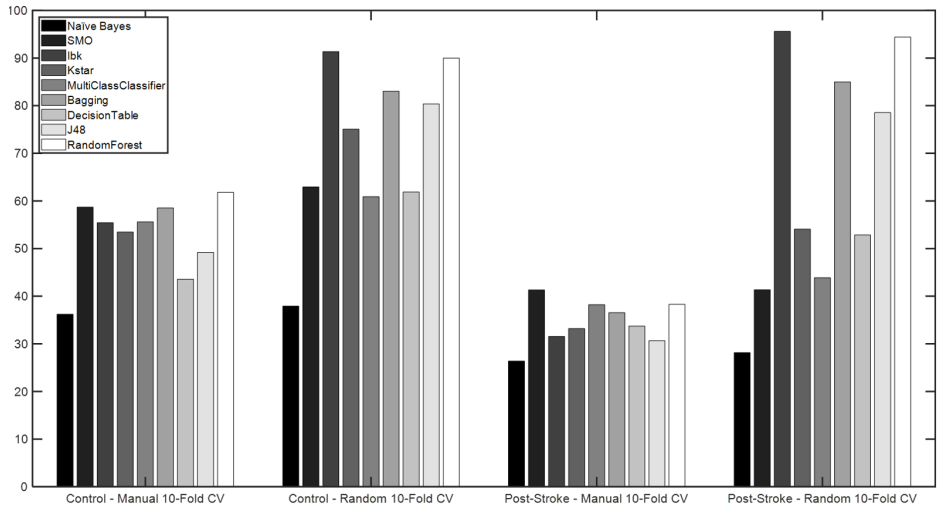


Figure 5.5: Correctly classified instances for persons Post-Stroke and Controls when the training data was randomly divided and when the data was separated manually, leaving two participants out to test the model and train the remaining participants.

While many features may be useful in human movement and activity recognition, our results suggest a subset may be sufficient to achieve the demonstrated classification accuracy. Choosing a selected number of features from the data set that were previously shown to be useful, did not significantly decrease classification accuracy. This suggests that, although there will likely be instances in which certain features are useful for a specific type of human activity monitoring, there are a subset of features which are more useful for general ADL classification.

Finally, when using overlapping two second windows to both generate and test the models, we demonstrated that it is possible to achieve high classification accuracy for both persons post-stroke and controls. Being aware that, with any level of overlap, a random 10-fold cross-validation is likely to train a model with data that is similar, if not identical, to the data that will be used to test the model, we also tested a non-random hand-selected 10-fold cross-validation. As expected, the classification accuracy decreased significantly for these user-independent models. However, given the ultimate application of this technique for real-time monitoring, it is not unreasonable to assume that a user-dependent model could be generated before real time monitoring begins. Given access to a participant's data (e.g., a calibration period, as used with current wearable technologies), it is possible to achieve classification accuracy over 90% from overlapping static two-second windows, suggesting the use of user-dependent models.

Although there is still work to be done to create a generalized model that can correctly classify task performance in post-stroke and control populations, this paper lays a groundwork for pursuing that goal. A number of variables and techniques that must be considered when developing a classifier are evaluated, and the variables and techniques necessary to achieve peak performance are ascertained. Our hope is that the efforts of this paper will be a progress point for the work being done to classify human motion using machine learning applications.

5.4 Data Exploration

5.4.1 Methods

As mentioned in **Section 5.1**, in addition to the supervised learning described in **Section 5.3.1**, several other learning methods and statistical models were used to evaluate or classify the time-series data that we had collected from our control and post-stroke participants. However, the results of all previous analyses never reached a level of accuracy sufficient for implementation in real time application. In order to determine why the classification accuracy of the five tasks remained low despite using many methods which had proven useful in other time series classification attempts previously, an exploration of the data was performed. First, participant data collected from the post-stroke participants had all been grouped together to this point, but the participants had a varied array of Fugl-Meyer (FM) scores indicating a large difference in the levels of impairment between the participants. Secondly, the analysis had also grouped left and right hand affected post-stroke participants. While the wrist and upper sensors of the affected limb were always labeled " Wrist 1" and "Upper 1" and the tasks were always separated so that tasks performed with the more affected limb were sorted from tasks performed with the less affected limb, the mirroring of the values could have an effect on the validity of the classifier. Finally, after grouping the data to determine whether those factors were affecting the outcome of the analysis, the raw data were analyzed to observe the underlying themes in each group and/or task type that were being missed in the previous approaches.

Fugl-Meyer Analysis

The Upper Extremity Fugl-Meyer (UEFM) was used to establish motor impairment based on a maximum score of 66 with a higher score indicating better motor function [75, 144, 194]. It is possible that a portion of the inability to accurately classify the data was because of the widely varied levels of impairment and function that are possible following the onset of stroke. A person post-stroke with a FM score of >= 60 might, when performing tasks meant to simulate activities of daily living (ADLs), generate IMU data more similar to a control participant than to a person post-stroke with a FM score of <= 10. Given that the

Fugl-Meyer is an assessment of impairment, it is possible that separating the participant data by their FM scores could increase the accuracy of classification.

For the analysis, the Fugl-Meyer score for each participant was plotted versus the participant labels (1-20 in the order the data was collected). From the figure (**Figure 5.6**), the outliers (empirically determined by FM score) were removed, leaving 17 participants in the "middle band." The middle band was then separated into upper (10 participants with FM scores $>= 30$) and lower (7 participants with FM scores < 30) portions to include the possibility of evaluating the more impaired individuals separately from the less impaired individuals.

Hemiparetic Stroke and Handedness

Following stroke, a number of individuals will demonstrate hemiparesis, or weakness of one side of an individual's body resulting in a more affected (paretic) limb and a less affected (non-paretic) limb [131]. Obviously, performance of tasks primarily involving the more affected limb will look significantly different that performance of tasks primarily involving the less affected limb. To account for these differences, the two tasks during which one hand was primarily active (unimanual and bimanual asymmetric tasks) were divided according to whether the more affected (or non-dominant) limb was active or the less affected (or dominant) limb was active. This means that for some participants, tasks classified as "Task 1" (or unimanual tasks with the dominant or less affected limb) were performed with the right hand and for others these tasks were performed with the left.

While the general amplitude and shape of these tasks as they are being performed should be similar between participants with different dominant and affected limbs, there is the possibility that the mirroring of IMU data was causing issues when it came to classifying the data. For each participant, it is known which limb is more dominant (controls) or more affected (post-stroke). Thus, the data were subdivided in this way to account for the differences that might occur based on the differences in limb function.

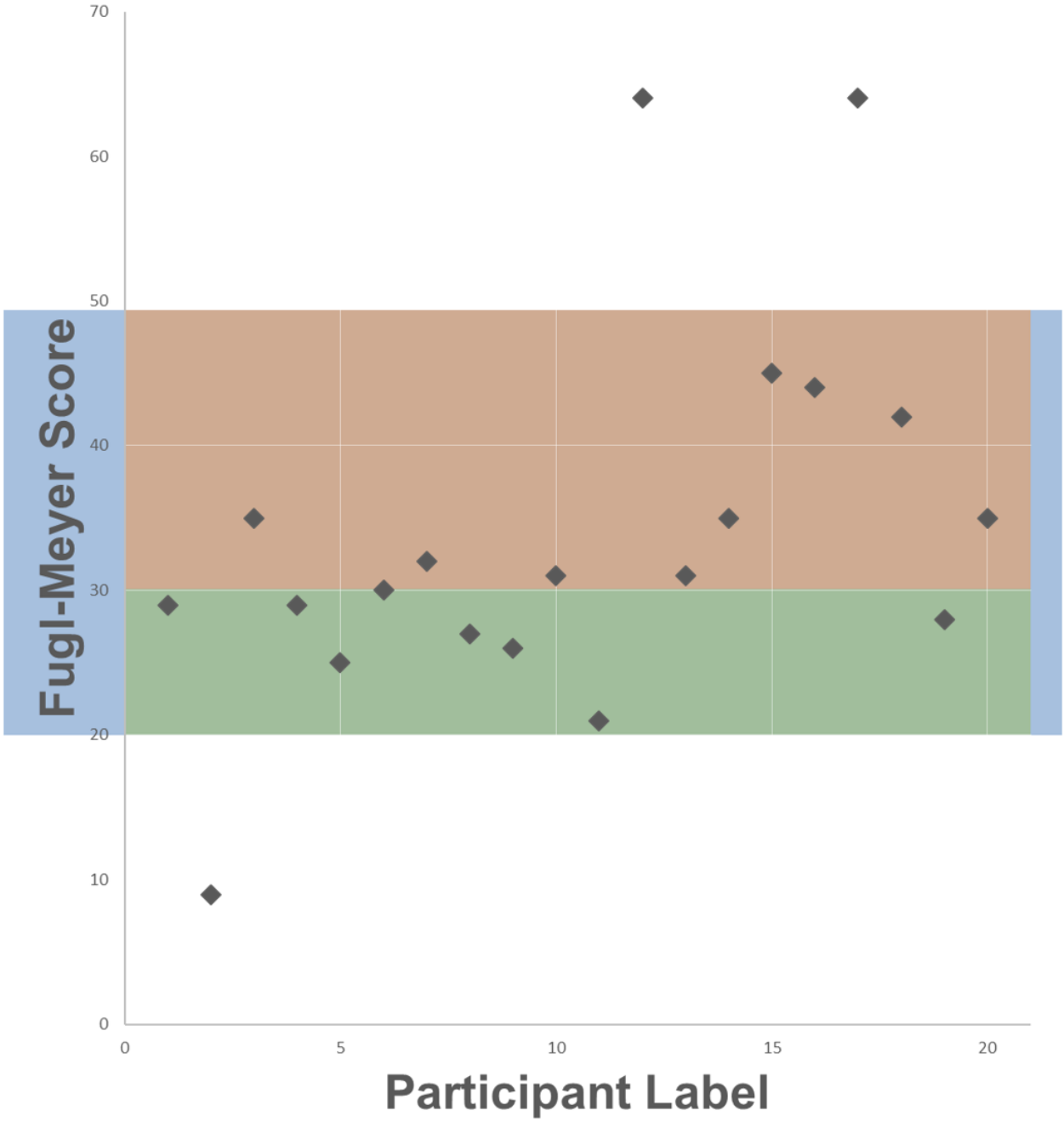


Figure 5.6: The Fugl-Meyer, which can be scored from 0 (worst performance) to 66 (best performance), is plotted versus the participant labels in the order the data was collected. The blue band represents the participants which are not outliers, the green band represents the lower portion of those participants which are not outliers, and the orange band represents the upper portion.

5.4.2 Results

All data were separated into two second windows with a 25% overlap and were tested in a Random Forest Classifier. The results of the analysis on the data after it had been separated into these different groups by Fugl-Meyer score is shown in **Table 5.1**.

The results of separating the participants by handedness are shown in **Table 5.2**. The handedness of the participants was accounted for in a few different ways: 1) by adding a single feature which represented whether the participant was left or right limb affected/dominant (**Binary L/R**); 2) by doubling the number of tasks, so that tasks performed by left hand affected/dominant participants were labelled separately from tasks performed by right hand affected/dominant people (**Distinct L/R**); 3) and finally by performing individual classification on participant groups based on hand affectedness/dominance (**Only L/Only R**). All analysis was initially performed on post-stroke participants as the difference in affected hand are more distinct than dominance, and when poor results were shown, only **Binary L/R** analysis was done for control participants to decrease computational time and efforts.

Finally, the results of the exploration of the raw data are shown in **Figure 5.7**. In the top row, we can see that participants post-stroke did engage the trunk more than control participants for all task types. In the second row, during bimanual symmetric tasks, both post-stroke and control participants demonstrate little use of the non-dominant or more affected wrists during bimanual symmetric tasks, but during unimanual and asymmetric tasks control participants engage the non-dominant limb more than post-stroke participants. From the third row, it is apparent that during all task performances, control participants engage the dominant/less affected limb more than post-stroke participants. The pattern from rows two and three repeats itself in rows four and five for the respective upper limbs. However, even given the distinctions that can be made from the chart, no statistically significant differences emerge in this raw data between the groups or tasks. This is because the between participant variability is high enough to hide the between task or between group differences.

Table 5.1: Average correctly classified instances when using only participants with certain FM scores and excluding those scores which were outliers: top 10 refers to the 10 participants whose FM scores were above 30, bottom 7 refers to the 7 participants whose scores were below 30, and middle 17 refers to a combination of the two aforementioned groups.

Data Used	**Average Classification Accuracy**
All	52.3%
Upper Band 10	53.9%
Lower Band 7	48.7%
Middle Band 17	52.9%

Table 5.2: Correctly classified instances post-stroke and control data when separated by handedness.

Data Used	**Accuracy (%)**	
	Post-Stroke	**Control**
All	42.6	-
Binary L/R	43.8	64.7
Distinct L/R	40.7	-
Only L	42.1	-
Only R	39.3	-

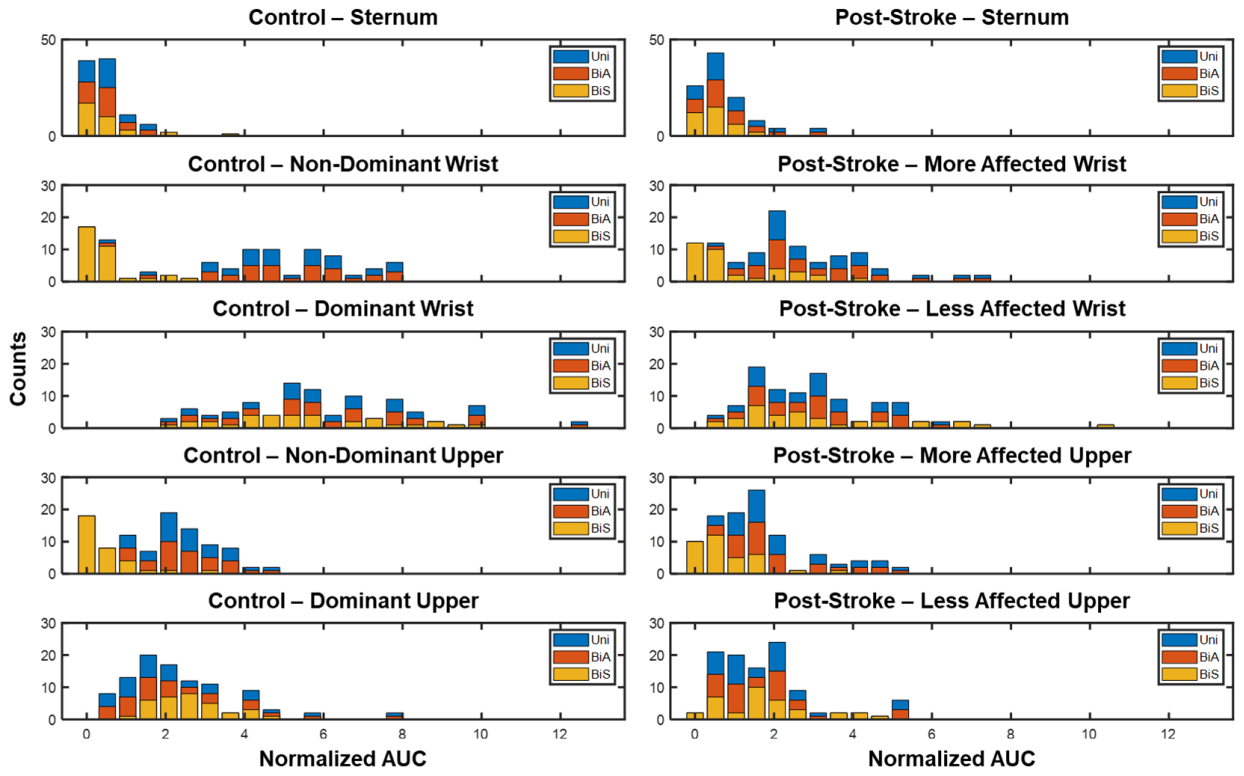


Figure 5.7: Exploration of the raw data. Data is shown in a histogram which demonstrates the normalized Area Under the Curve (AUC) for a given segment and task. The left column shows control participants and the right post-stroke; the rows represent the sternum, non-dominant/more affected wrist, the dominant/less affected wrist, non-dominant/more affected upper, and the dominant/less affected upper respectively. The color of the bars demonstrate the task being performed: blue for unimanual, red for bimanual asymmetric, and yellow for bimanual symmetric.

In **Section 5.3.1**, several features were extracted from the data to prevent between participant variability from hiding differences between participant groups or task performance when looking at a given extracted feature. The expectation is that extracting multiple features would better help to differentiate between the tasks and groups. However, when we began to explore the original data by extracting features and looking for systemic differences between groups and tasks, we found that for each feature, between participant variability was hiding any systemic differences. This between participant variability causes our data to present as chaotic.

The study of chaotic systems dates back to the late 19th century and covers topics such as the weather, stock market, and brain activity [147]. A chaotic systems is defined as having aperiodic long-termed behavior that exhibits sensitive dependence on initial conditions [33]. While it is possible to record time series data of one or a few dynamical variables of a system, the equations that govern the system are unknown, and often unknowable, until deeper data analysis approaches are employed [147]. Human motion is often analyzed using nonlinear or chaotic systems [23, 201, 236] because the disparate ways that different humans originate motion can lead to highly varied data. From this point on, when we refer to the data we collected as chaotic, we are referring to the between participant variability that occurs as a result of various factors including origins of movements or submovements.

5.5 Discussion

In this chapter, we discussed how many different machine learning methods were applied to the data in an attempt to classify between the post-stroke and control groups as well as between the unimanual, bimanual symmetric and bimanual asymmetric task types. The importance of distinguishing between tasks types is discussed in **Section 3.5** where analysis of the literature demonstrated that compensation is almost always defined in the context of a segment and task being performed when compensation is measured.

To begin classifying task types between groups, first supervised learning methods were applied. These methods are commonly used for classification, so a number of machine learning classifiers, features, window sizes, and percent overlaps were analyzed to determine

the effectiveness of these techniques for classifying time-series, human-motion data. From this analysis, we learned that it is possible to create a classification model that can accurately classify data from a participant given that data from that participant is used to train the model. However, in exploring the generation of a more generalized model, the results were not as promising. We did learn that decreasing the number of features to a few specialized features decreased computational time but did not significantly decrease the results of the classifier. We also learned that because of the speed with which some tasks are performed, both in the lab and ambient setting, that changing the window size between one second, two seconds, and full time required to complete the task did not significantly affect the outcome. These results are promising as both decreased computational load and fixed window sizes are important features for real-time analysis of time series data.

We attempted several other classification techniques, but we never achieved an acceptable level of classification accuracy. To determine why our accuracy remained low, we began an exploration of the raw data. This exploration showed that the between participant variability was high enough to hide any between group and between task variability in the data. This variability was primarily demonstrated in the representative features that were extracted from the raw data, which was done to decrease computational load. For the majority of the analysis, derived features were used as inputs for the models, and even in the instances where the raw data were clustered, the input for the learning models were still the outputs of the clustering that was performed. Through this analysis, we learned that adding more features does not continue to improve accuracy, and that static windows can be used to accurately classify gestures. To implement this, in the next chapter, we explore a methodology which addresses the problem of oversampling by synthetically augmenting underrepresented datasets and removing overlapping (or correlated) data points. We then use the raw data [38] and a minimal number of features in some neural networks, which are capable of quickly classifying data in real time.

Chapter 6

Classifying Unimanual and Bimanual Upper Extremity Tasks in Individuals Post-Stroke

6.1 Introduction

As previously mention, stroke is the leading cause of serious long-term disability in the US [267]. To maintain function and independence following stroke, a survivor will often employ compensatory strategies [137, 265, 174] which can have negative long-term effects [241, 177]. These behaviors typically develop in the ambient setting [177], so it has become necessary for clinicians to be able to remotely monitor patients post-stroke [257, 243, 199].

In **Chapter 3**, we determined that it is necessary to identify a task being performed before compensation can be quantified [180]. In **Chapter 4**, this phenomenon was investigated using a selected set of upper extremity (UE) tasks. **Chapter 4** demonstrated that wearable sensor data are sensitive to differences in task types as well as the differences in task performance by post-stroke and control groups. **Chapter 5** also showed that a select set of features extracted from overlapping data windows can be used to train a supervised machine learning model to distinguish between task types. However, given the complexity

and variability of UE motion, achieving high accuracy required classification techniques trained and evaluated on the same datasets (i.e., user–dependent models). The goal of the current work is to determine if other machine learning approaches are capable of creating a generalized, universally applicable (i.e., user–independent) model capable of distinguishing between task types for post-stroke and control individuals using windowed motion data.

Machine learning methods have been used in many applications for people post-stroke. Health outcomes [89], therapy stage [181], and Fugl-Meyer Assessment (FMA) scores [250] have been predicted using supervised machine learning techniques (random forest, linear regression, and Support Vectors Machines (SVM)) and deep neural networks. Hand gestures have been classified using supervised techniques such as linear discriminant analysis (LDA), SVM, and k-Nearest Neighbors (KNN) [107, 140], and gait and walking conditions have been characterized by SVM and neural networks [109, 132]. To date, machine learning methods have not been extensively used to classify upper extremity (UE) gestures. This is likely because time series data that present with seasonal cycles combined with linear trends or other randomness, such as UE motion data, often appear chaotic and can be difficult to classify using a time series model. However, the echo state neural network (ESNN) is a recurrent neural network that uses a loosely connected hidden layer (known as a reservoir) and works well with chaotic data [193, 271].

Other data processing techniques, such as LDA [152, 151], quadratic discriminant analysis (QDA) [229], Naive Bayes (Gaussian and Kernel Density) [30, 286], decision trees [249], canopy clustering [172, 142], centroid clustering [223], and principal component analysis [101, 171, 95] have been used in conjunction with neural networks to improve results of classification or analysis. Some clustering and dimensionality reduction techniques have been shown to improve the accuracy of neural network classifiers [191]. To investigate the utility of this approach with UE data we use a multi-layer perceptron (MLP) artificial neural network (ANN) to establish baseline accuracy. We then investigate the performance of neural network–based task classification, relative to the previously employed machine learning methods. After establishing a baseline, we evaluated a preprocessing technique to examine its effects on the classifier accuracy. However, because this technique was not included in the final analysis, we moved the discussion of this technique to **Appendix B**

to improve the narrative flow of this document. We then test an ESNN to evaluate model accuracy given complex, highly variable, windowed time-series data from post-stroke and control groups performing UE tasks.

6.2 Methods

6.2.1 Design

Individuals post stroke and healthy-age matched controls were recruited from two clinical sites: Columbia University Irving Medical Center/Teachers College, Columbia University and Chapman University. In this cross-sectional study, participants attended a single session where they were asked to perform a series of tasks, while both seated and standing, that are representative of the types of tasks that an individual would be performing in an ambient setting. These tasks are described in **Section 5.2.2** and consisted of: UL, UM, BAL, BAM, and BS tasks. This study was approved by the Institutional Review Boards at Columbia University Irving Medical Center, Teachers College, Columbia University and Chapman University. All participants provided written informed consent.

6.2.2 Procedure

Participants wore five APDM Opal inertial measurement units (IMUs) on the sternum, both wrists, and both upper arms during task performance. Each sensor recorded three sensor modalities (accelerometer, angular rate of change, and magnetic field strength) on three axes (x, y, and z) at 128 Hz (www.apdm.com). Participants were also filmed to obtain ground-truth task performance times. These data were filtered using a Butterworth bandpass filter with cutoff frequencies of 0.1 Hz–2 Hz (this filter was determined to be appropriate for signal power of volitional movement common to activities of daily living, (ADLs) [188]). Finally, data were detrended to remove effects of drift. Following task completion, video data were reviewed by a clinician who manually determined start and stop times for each task resulting in a dataset consisting of twelve tasks (2 UL, 2 UM, 2 BAL, 2 BAM, 4 BS), 40 participants

(20 post-stroke and 20 control), and 45 signals (5 sensors×3 modalities×3 axes). Each task was then divided into two second windows, with 50% overlap.

6.2.3 Feature Extraction

For each window, data were downsampled to a sampling rate of 64 Hz (from 128 Hz) to increase computational speed. We then extracted various features useful for differentiating and classifying human motion [29] (described in **Section 5.3.1**: Hand-Picked Feature Extraction) including: area under the curve (AUC); root mean square (RMS); signal magnitude area (SMA); signal vector magnitude (SMV); energy; entropy; fast Fourier transform peak (FFTP); mean; and standard deviation (SD). Each two second window resulted in a matrix of 45 values (5 locations × 3 modalities × 3 axes) by 128 (two seconds × 64 Hz).

Beginning with the second data point ($2/64^{\text{th}}$ seconds into the window), features were extracted from subvectors of the raw data that increased in size until the final extracted feature was derived using raw data from the entire window. For features such as mean, methods like this are already commonly employed (e.g. cumulative moving average [1]). However, in our case, each new feature is extracted from the previous subvector of raw data appended by the latest datum point. For the first instance of data in the window ($1/64^{\text{th}}$ of a second into the window) there are insufficient data points to create a vector from which features can be extracted. To counteract this and ensure that feature vectors have the same length as the raw data, a new data point was approximated from the raw data. The new data point is assumed to happen at time $t_0 = 0$s and the data point itself is approximated by the formula below where d_1 and d_2 are the data points that are found at time $t_1 = 1/64^{\text{th}}$ (s) and $t_2 = 2/64^{\text{th}}$ (s) respectively:

$$d_0 = d_1 - [d_2 - d_1] = 2 \times d_1 - d_2$$

This formula assumes equal spacing between data points and allows for the generation of a feature vector of the same length as raw data. Table 6.1 illustrates how values are generated for **one** feature using this method. This approach was used to ensure that all features were

Table 6.1: Data synthesis approach

n	1	2	...	128
Feature	f_1	f_2	...	f_{128}
Time	$[t_0, t_1]$	$[t_1, t_2]$	...	$[t_1, t_2, ..., t_{128}]$
Data	$[d_0, d_1]$	$[d_1, d_2]$	...	$[d_1, d_2, ..., d_{128}]$

appropriately formatted as inputs for MLP and ESNN models (creating an input matrix where each row has the same number of columns).

6.2.4 Data Preprocessing and Shaping

Whether establishing a baseline with the multi-layer perceptron or using the echo state neural network, the data need to be normalized and shaped to fit the classifier. Each feature was standardized (subtracting the appropriate means and dividing by standard deviation for each modality) to account for relative differences in magnitude. Because the windowing technique was used to generate input data, tasks that took longer to perform (e.g., folding a towel) were overrepresented while other tasks (e.g., reaching for a spoon) were underrepresented. To account for this, participants were randomly combined in groups and their data were balanced using the Synthetic Minority Oversampling Technique (SMOTE) [39] combined with Edited Nearest Neighbors (ENN) [25]. Both techniques are used to synthetically balance unbalanced datasets. They also removed data points at the overlap between classes, resulting in better accuracy and reduced computational costs. Sixty percent of the data from each participant were randomly selected to train the classifier and the other forty percent were held for validation and testing.

Finally data were reshaped into a three-dimensional matrix where each row corresponds to a single feature for a single sensor, each column represents a time step, and each layer represents a modality. Input data were of dimension $15N$ (where N is the number of included features) by 128 (2 second window by 64 Hz sampling frequency) by 3 modalities.

6.2.5 Model Structure

Data were classified using two different neural networks; a simple MLP and an ESNN. The MLP was chosen because of its relatively low computational cost and its frequent use for nonlinear time series prediction [155, 16, 242, 103]. Our selected MLP has an input layer, a hidden layer consisting of a flattened layer and 5 dense layers (4 with the rectified linear unit activation and a final dense layer with a number of nodes corresponding to the number of task types (5) with a SoftMax activation), and an output layer. The number of dense layers in

the hidden layer and the number of neurons per dense layer were empirically determined by systematically iterating numbers of layers and neurons. The MLP was then compiled using the Adam optimizer which is computationally efficient and works for classification problems with large amounts of data [118]. The loss function used was categorical cross-entropy which computes the cross-entropy loss between the labels and predictions. The model was then fit to the training set and classification performance was improved by minimizing the cross-validation error over several iterations.

The hyper-parameters of the ESNN were also empirically determined to maximize classification accuracy. Using a local, random search, the selected hyper-parameter values were similar to those utilized in [234] (Table 6.2).

6.3 Results

To obtain baseline model performance, the MLP was first evaluated using preprocessed raw data only (RAW). We then used a literature-based approach to adding model features, assessing a model using preprocessed raw data and the extracted area under the curve features (RAW+AUC). Next, a third feature was added (RAW+AUC+f_3). The third feature was one of the list from Section 6.2.3. The features for which highest accuracy was achieved (RMS, SMA, SVM, FFTP) were then used in combination with RAW+AUC data, as well as each other, to evaluate the model for both post-stroke and control participants. Only the highest performing features were used to increase computational efficiency and decrease time spent on analysis. Finally, a fourth feature was added (RAW+AUC+f_3+f_4). No additional features were added to the input dataset because doing so increases the chance of correlated features, which can lead to model over-fitting and decreased validation and testing accuracy [276, 274]. Results are depicted in **Figure 6.1**.

We used the Mann-Whitney U test to determine if adding the fourth feature improved accuracy (when compared to RAW+AUC+f_3). For control participants, continuing to add features after RAW+AUC did not increase model accuracy and as the number of features increased from three to four, testing accuracy significantly decreased (accuracy $p = 0.071$; validation accuracy $p = 0.439$; testing accuracy $p <= 0.003$). The highest testing accuracy

Table 6.2: MLP Hyper-parameters

Hyper-parameters of the Reservoir	
Size of the Reservoir	495
Largest Eigenvalue of the Reservoir	0.53
Leakage in the Reservoir State Update	0.66
% of Nonzero Connections in the Reservoir	25
Scaling of the Input Weights	0.0878
Noise in the Reservoir State Update	0.01
Transient States to be Dropped	5
If True, Use Bidirectional Reservoir	False
Use Reservoir with Circle Topology	False
Dimensionality Reduction Hyper-parameters	
Dimensionality Reduction Method	PCA
Number of Resulting Dimension After PCA	168
Multivariate Time Series (MTS) Representation	
MTS Representation	`reservoir`
Parameter of the Ridge Regression	11
Type of Readout (Linear, SVM, MLP)	
Readout used for Classification	`linear`
Linear Readout Hyper-parameters	
Reg. of the Ridge Regression Readout	4
SVM Readout Hyper-parameters	
Bandwidth of the RBF Kernel	0.0055
Reg. of the SVM Hyperplane	5.5
MLP Readout Hyper-parameters	
Neurons in each MLP Layer	9
Number of Epochs	2000
Weight of the L2 Regularization	0.001
Type of Activation Function	`tanh`

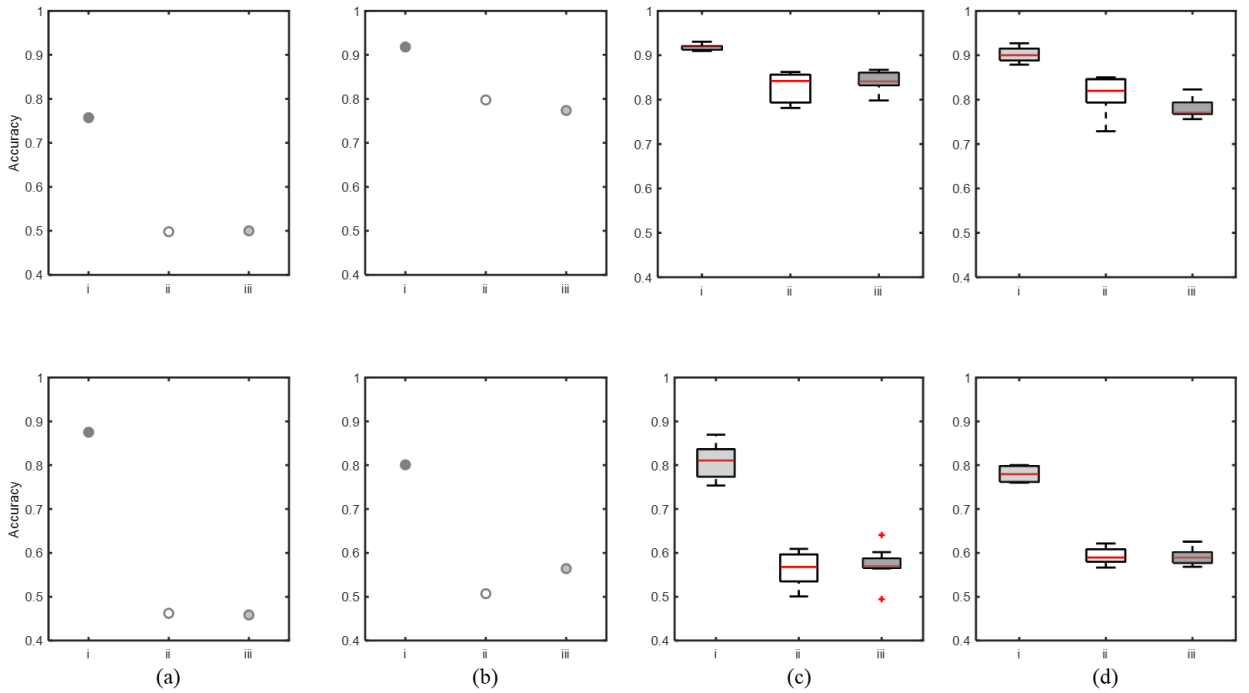


Figure 6.1: Accuracy (i), validation accuracy (ii), and test accuracy (iii) of the MLP classifier when using: (a) raw data; (b) raw data and the derived AUC feature; (c) raw data, AUC, and one other derived feature; or (d) raw data, AUC, and two other derived features for both the Control (**row 1**) and Post-Stroke (**row 2**) groups

for controls (86.7%) was achieved using the RAW+AUC+SMA features. For participants post-stroke, although the overall average testing and validation accuracy continued to increase as more features were added, improvements were not significant (accuracy $p = 0.156$; validation accuracy $p = 0.156$; testing accuracy $p = 0.197$). The highest testing accuracy for participants post-stroke (64%) used RAW+AUC+RMS features.

The ESNN was evaluated using all three readouts used for classification. For control participants, testing accuracy was 91.3%, 86.5%, and 87.1% (when using the linear, support vector machine, and multi-layer perceptron readout, respectively). For participants post-stroke, the testing accuracy was 80.3%, 74.1%, and 76.7% (when using the linear, support vector machine, and multi-layer perceptron readout, respectively). Finally, both the post-stroke and control data were used to train a classifier, resulting in ten tasks corresponding the five task types for each of the two groups. If the performance of the age-matched controls is designated as how a task would be performed before the onset of stroke (or non-compensatory), then the next steps in being able classify compensatory behavior include differentiating between non-compensatory (control) and compensatory (post-stroke) performances of a task. The results of that analysis for both groups combined show that the testing accuracy was 81.4%, 73.0%, and 75.9% (when using the linear, support vector machine, and multi-layer perceptron readout, respectively).

6.4 Discussion

The purpose of this analysis was to determine if a patient-independent model capable of classifying noisy, chaotic, human-derived, time series data could be created through the application of novel classification approaches (including the multi-layer perceptron and echo state neural network). This new approach improves on the accuracy of the results we achieved in prior analyses(**Chapter 5**, [180]). Our previous analyses demonstrated data collected from each participant were necessary to train accurate classification models prior to real time use (i.e., user-dependent models). Use of the MLP and ESNN to create generalized models will bypass the need to collect and add participant data to the model and facilitate the use of real time, user independent models.

Results of our MLP model indicated that, although the validation and testing accuracy continued to improve for participants post-stroke as feature were added, adding features did not significantly increase the classifier accuracy. Significant decreases in testing accuracy with increased number of features for control participants was likely due to model over-fitting. Given the risk of over-fitting and the increased computational cost of adding features, we conclude that it is best to let the neural network extract relevant features.

When using the ESNN, classification accuracy greatly increased relative to MLP performance. This analysis has shown that a complex RNN such as the echo state neural network can create a generalized model that can be used in the real-time classification of tasks. This model can be designed to work for both control participants and participants post-stroke using minimally intrusive sensing techniques that do not impede the function of a participant.

As mentioned in the introduction, the first step in autonomous recognition and reversal of compensatory movement strategies is the capability to differentiate between the types of tasks that people post-stroke perform in the ambient setting. This classification can then be further used to identify, address, and decrease the use of compensatory behaviors that can lead to a decreased quality of life or lifespan. In the future, work can be done to improve the results of the analysis using different preprocessing techniques to limit the likelihood of overfitting, to determine the impact that data from each sensor location has on the outcome of the models, and to implement the system in real-time to validate the models on a new group of participants.

6.5 Acknowledgement

The authors thank Dr. Susan Duff, EdD, PT, OT/L, CHT, for her ongoing collaboration in collection, analysis, and contextualization of data. This work was supported by NSF Award #2054191.

Chapter 7

Wearable Sensors for Reducing Compensation in Persons Post-Stroke at Home

7.1 Introduction

Following stroke, impairments such as hemiparesis [138, 245, 294] and decreased interlimb coordination [8, 237, 216, 235, 119] are common and can lead to decreased independence [128], quality of life, and lifespan [82, 130, 280]. To regain function and independence, persons post-stroke will rely on compensatory strategies to help perform activities of daily living [137, 265, 174]. However, maladaptive compensatory strategies can lead to long-term negative effects [241, 177, 35, 146, 6, 174]. These compensatory movement strategies evolve in the ambient setting [177], and recently, research has focused on autonomously detecting and reversing compensatory behaviors [257, 243, 199].

Learned Non-use

One common compensatory behavior is non-use of the affected limb. Continued disuse of the more affected limb following a stroke or other neurodegenerative disorder can lead to a phenomenon known as learned non-use. Because of the plasticity of the brain, when the

paretic limb is not being used, the brain reorganizes neural pathways which had previously been dedicated to the use of the more affected limb. Given that other body segments are now being required to perform more complex tasks than they had been previously, more resources are dedicated to the functioning body segments. This, in turn, leads to further impairment of the more affected limb, which then leads to further disuse of the more affected limb, and the cycle continues. This results in a person post-stroke not only training himself/herself out of regularly using the more affected limb, but also weakening their ability to do so if desired [241, 189, 238].

Constraint Induced Movement Therapy

Constraint induced movement therapy (CIMT) is the most prevalent method used for reversing non-use of the affected limb. During CIMT, the less affected limb is constrained, typically by use of a mitt, forcing a person post-stroke to use the more affected limb [127]. A person post-stroke is required to constrain use of the less affected limb in this way for approximately 90% of waking hours, and he/she is encouraged to participate in intensive practice of the more affected limb to increase task specific performance. As such, CIMT has been proven to increase recovery rates, improve rehabilitation, and reduce physical impairment and rewrite the brain, reversing the effects of learned non-use [60, 143, 58]. While CIMT has obvious benefits, it is also responsible for reducing functional independence (as both hands can not be used to complete a task) and might be responsible for reducing use of the less affected limb [282]. Researchers have modified CIMT (mCIMT) dose, timing, and composition, and have demonstrated that mCIMT has positive effects on motor function [127]. However, as a participant's functional independence is still severely limited, there is a need for a less invasive rehabilitative protocol for persons post-stroke.

Feedback Induced Movement Therapy

One such noninvasive approach is monitoring patient movement via minimally intrusive, body-worn sensors. These sensors have a low impact on the daily routine of a person post-stroke, but they have been proven capable of measuring features such as nonuse and providing feedback to participants when specific compensatory strategies are detected. This Feedback

Induced Movement Therapy (FIMT) operates by using targeted messaging to incentivize the use of the more affected limb rather than forcing use, which can be a method for revealing a participant's inherent movement capability even if it is less preferred by the patient following stroke [18, 149, 167]. The difference between FIMT and CIMT for patients post-stroke, including those with a high reliance on the less affected limb, is that, where CIMT can impede function by limiting movement, FIMT promotes use of the more affected limb without limiting function.

Accelerometry in Stroke Rehabilitation

One type of sensor useful for FIMT in stroke rehabilitation is the accelerometer [273]. Using accelerometers, researchers have been able to reliably monitor and measure paretic limb use and to index upper extremity function in an ambient setting [255, 257, 260]. The use of these sensor based techniques provide daily information to the patient and are helpful in providing the clinician with insight into the patient's daily life as they plan interventions [149]. Sensor based feedback from accelerometers has been shown to increase self-efficacy and physical activity in patients post-stroke in the clinical or research setting, albeit in the lower extremity [78, 112, 120, 162]. Use of these sensors has also been shown to increase a clinician's knowledge of a patient's mobility related activity [65]. Given that, using accelerometer based sensing techniques, a clinician can collect an individual's in-home movement data, it is now possible to generate personalized rehabilitation strategies tailored to each individual patient [3].

However, just having knowledge of a patient's activity is often not enough for a clinician to design a rehabilitative routine for each specific participant. The accelerometer can provide a reliable and valid index of upper extremity motor activity in patients with hemiparesis throughout the stroke recovery process [257]. These sensors can also measure the ratio with which each arm is used by demonstrating that people use the paretic upper limb significantly less than controls after stroke [131, 221]. In the chronic stage, accelerometers have been used to demonstrate that use of the less affected limb increases [210], particularly after therapy like CIMT [195, 256, 240]. The above examples fail to differentiate between general movement and task related movement. As in-home rehabilitative techniques become less invasive and

more practical for everyday use, such as body-worn sensors, new techniques need to be developed to provide insight into upper extremity function.

Previous Work

In **Chapter 4**, work was done to determine if it was possible to quantify non-use and predict clinical measures using sensor data from participants post-stroke. Data collection was performed on twenty participants post-stroke and twenty age matched controls. Participants were asked to perform twelve different tasks which are described in **Section 5.2.2** and consisted of: 2 UL, 2 UM, 2 BAL, 2 BAM, and 4 BS tasks. Data collection was done using five IMU, and is described in **Section 4.2**. In **Chapter 6**, using the information collected from the sensors during these simulated functional tasks, we were able to develop a machine learning model that could differentiate between the five aforementioned task types and the two groups (post-stroke and control) [180, 179].

The primary purpose of this study is to develop and validate a non-intrusive, sensor based system for providing Feedback Induced Movement Therapy to a control participant in an ambient setting. In **Chapter 6**, we developed an autonomous, machine learning approach for differentiating between tasks and detecting compensatory behaviors [180, 179]. This study will focus on the development of the autonomous feedback delivery method via a project specific cell phone application (RehApp), and will validate the system in an ambient setting; testing feedback at a chosen interval to determine amount of intervention engagement and intervention fatigue [185]. Because of time constraints, the system will not be tested on a participant post-stroke, but will instead be validated on a single, non-age matched participant to confirm it works as expected.

7.2 Methods

"Triple Play" is a feedback development guideline introduced by Wulf et. al. [290] which suggests that participants will respond far more positively when provided feedback: enhances their expectancy of results, provides them with autonomy, and externalizes their focus [290, 62, 49, 289]. Enhanced Expectancies (EE) is a motivational variable that has been shown

to promote learning. Examples of EE include: a participant receiving feedback only after performing well during a trial [13, 44, 52, 217]; being told, either truthfully or falsely, that his or her performance is better than the performance of peers [11, 139]; being told that peers typically perform well on the given task [288]; or increasing the participant's perception of success during training [43, 288, 196, 251]. Autonomy (AS) is another motivational variable that is important for optimal learning. Allowing participants to control various aspects of practice in motor learning (delivery of feedback [104], use of assistive devices [84], extent of practice [203], and frequency of skill demonstration [291]) has been shown to result in more effective learning. The final variable, External Focus (EF), is attentional and directs the participant's focus to the intended movement effect. This has been shown to enhance learning when compared to feedback that has an internal focus on body movements or not instructed focus [150, 164, 287]. Use of these positive reinforcement strategies has been shown to be key to optimizing motor recovery following stroke. All of the feedback to be provided to participants in the trial was developed according to the guidelines laid out in "Triple Play." The feedback, as well as the conditions for providing feedback are shown in **Table 7.1**.

In addition to the content of the feedback provided, the frequency of feedback, or number of times feedback is given in a session, is important to the efficacy of the system. Recent work, examining Just-In-Time Adaptive Intervention (JITAI), which is an intervention design that aims "to provide the right type (or amount) of support, at the right time, while eliminating support provision that is not beneficial," suggests that intervention adherence and retention are dependent on two important factors: engagement and fatigue [185]. Intervention engagement is defined as a "state of motivational commitment or investment in the client role over the treatment process" [117], and intervention fatigue is defined as a "state of emotional or cognitive weariness associated with intervention engagement" [185, 87]. If it can be assumed that patients participating the trial will demonstrate intervention engagement, then testing needs to be done to measure and mitigate intervention fatigue by adjusting the frequency of the feedback that is provided.

As mentioned in **Chapter 6**, we developed a model capable of distinguishing between task types and between groups with 81.4% accuracy [179]. The classifier is modular, and

Table 7.1: Feedback Conceptualization

Compared Period	Feedback Condition	Feedback
Current Period	Overall Healthy Use > Overall Unhealthy Use	You demonstrated healthy activity in the last time period.
	Healthy Bimanual Use > Unhealthy Bimanual Use	You demonstrated healthy two-handed activity in the last time period.
	Healthy Unimanual Use > Unhealthy Unimanual Use	You demonstrated healthy one-handed activity in the last time period.
	Healthy Bimanual Symmetric Use > Unhealthy Bimanual Symmetric Use	Your two-handed activity in the last time period was great.
	Healthy use of the More Affected Limb > Unhealthy use of the More Affected Limb during Bimanual Asymmetric tasks	During two-handed activity in the last time period, your affected arm use was great.
	Healthy use of the More Affected Limb > Unhealthy use of the More Affected Limb during Unimanual tasks	During one-handed activity in the last time period, your affected arm use was great.
	Use of the More Affected Limb > Use of the Less Affected Limb during Bimanual Asymmetric tasks	During two-handed activity in the last time period, you used your affected hand more than your unaffected hand.
	Use of the More Affected Limb > Use of the Less Affected Limb during Unimanual tasks	During one-handed activity in the last time period, you used your affected hand more than your unaffected hand.
Previous Period	Overall Healthy Use compared to Overall Unhealthy Use in Current Period > Previous Period	You demonstrated healthier activity than the previous period.
	Healthy Bimanual Use compared to Unhealthy Bimanual Use in Current Period > Previous Period	You demonstrated healthier two-handed activity than the previous period.
	Healthy Unimanual Use compared to Unhealthy Unimanual Use in Current Period > Previous Period	You demonstrated healthier one-handed activity than the previous period.
	Healthy Bimanual Symmetric Use compared to Unhealthy Bimanual Symmetric Use in Current Period > Previous Period	Your two-handed activity in this time period was better than the previous period.
	Healthy Use of the More Affected Limb compared to Unhealthy Use of the More Affected Limb during Bimanual Asymmetric Tasks in Current Period > Previous Period	During two-handed activity in this time period, your affected hand use was better than the previous period.
	Healthy Use of the More Affected Limb compared to Unhealthy Use of the More Affected Limb during Unimanual Tasks in Current Period > Previous Period	During one-handed activity in this time period, your affected hand use was better than the previous period.
	Use of the More Affected Limb compared to Use of the Less Affected Limb during Bimanual Asymmetric Tasks in Current Period > Previous Period	During two-handed activity, you used your affected hand more often this time period than the previous period.

Continued on next page

Table 7.1 – Continued

Compared Period	Feedback Condition	Feedback
Previous Period	Use of the More Affected Limb compared to Use of the Less Affected Limb during Unimanual Tasks in Current Period > Previous Period	During one-handed activity, you used your affected hand more often this time period than the previous period.
Current Period	Overall Healthy Use ≈ Overall Unhealthy Use	You demonstrated pretty healthy activity in the last time period.
	Healthy Bimanual Use ≈ Unhealthy Bimanual Use	You demonstrated pretty healthy two-handed activity in the last time period.
	Healthy Unimanual Use ≈ Unhealthy Unimanual Use	You demonstrated pretty healthy one-handed activity in the last time period.
	Healthy Bimanual Symmetric Use ≈ Unhealthy Bimanual Symmetric Use	Your two-handed activity in the last time period was pretty good.
	Healthy use of the More Affected Limb ≈ Unhealthy use of the More Affected Limb during Bimanual Asymmetric tasks	During two-handed activity in the last time period, your affected arm use was pretty good.
	Healthy use of the More Affected Limb ≈ Unhealthy use of the More Affected Limb during Unimanual tasks	During one-handed activity in the last time period, your affected arm use was pretty good.
	Use of the More Affected Limb ≈ Use of the Less Affected Limb during Bimanual Asymmetric tasks	During two-handed activity in the last time period, you used your affected hand almost as much as your unaffected hand.
	Use of the More Affected Limb ≈ Use of the Less Affected Limb during Unimanual tasks	During one-handed activity in the last time period, you used your affected hand almost as much as your unaffected hand.

can work with a number of different groups, tasks, and window sizes, but for the purposes this study, the data will, in real-time, be divided into individual two-second windows with a 50% (one second) overlap. Each window will then be classified by task type (unimanual (1): more affected, unimanual (2): less affected, bimanual asymmetric (3): more affected, bimanual asymmetric (4): less affected, and bimanual symmetric (5)) and by group (control and post-stroke). For the purposes of this study, task performances which are assigned to the "Control" group are considered non-compensatory (having been labeled as the way a task would have been performed before the onset of stroke) and those assigned to the "Post-Stroke" group are considered unhealthy/compensatory.

In this pilot trial, feedback will be delivered after every 15 minute period during a four hour long training session. After the feedback period has elapsed, the results of the classifier for each window over the period will be analyzed and the feedback will be delivered based on logic designed to provide positive feedback given any scenario (shown in **Table 7.1**). All analysis of data and delivery of feedback is to be completed through a cell phone application (RehApp) which was developed specifically for this research. The pilot participant will receive a notification containing the selected feedback at the end of each feedback period. The participant will also be able use the application to observe personal feedback and performance history.

7.2.1 Population

An individual dwelling within the community was recruited according to the standards of the University of Tennessee Internal Review Board. The participant recruited for the trial was age 27 and was be able to follow simple instructions and understand and comply with the requirements of the study.

7.2.2 Design

In this proof of concept, the participant was asked to wear the sensors for four hours. The sensors chosen were Shimmer Wearable Sensor Technology's "Wearable IMU [225]." As mentioned in **Chapter 2**, one of the goals for this project was to use wearable off the

shelf sensors to make this system more readily available for disadvantaged groups, such as persons post-stroke. Before donning the sensors, each sensors was configured in the ConsensysPRO Software [224] to ensure the accelerometer, angular rate of change, and magnetic field strength were being recorded at the correct frequency. As in the initial data collection detailed in **Chapter 4**, the participant wore the sensors on the sternum, both wrists, and both upper arms. While wearing the sensors, the participant was asked to perform normal activities of daily living, including folding laundry (BS), baking cookies (BAM/BAL), and building a puzzle with just one hand at a time (UM/UL). These ADLs were chosen for their similarity to the tasks that were described in **Chapter 4** and used to build the ESNN in **Chapter 6**.

An application was designed by an external expert consultant, Dr. Aravind Sundararajin, to connect to the sensors and provide feedback. The application was designed in Android Studio using a base Java/Android API created by Shimmer [226]. The API is designed to support connecting sensors and streaming data. The screen for connecting and configuring the sensors and beginning the data stream is shown in **Figure 7.1(a)**. The application was then further updated to package the recorded data and send it to a server where each package is classified using the echo state neural network developed in **Chapter 6**. The server stores the output of the classifier for each package until the frequency period has elapsed. When the frequency period is elapsed, the server uses the logic shown in **Table 7.1** to push feedback to the participant via cellphone notification. At any time during the trial period, the participant was able to request feedback, and the appropriate feedback for the frequency period would be sent. The home screen where feedback can be requested is shown in **Figure 7.1(b)** and the navigation options shown are described in the caption below.

7.3 Results

The purpose of this proof of concept was to validate the system in the ambient setting in real-time. After wearing the sensors for four hours, the stream was stopped and the results were compiled. A total of 4998 overlapping two-second windows were classified. The results

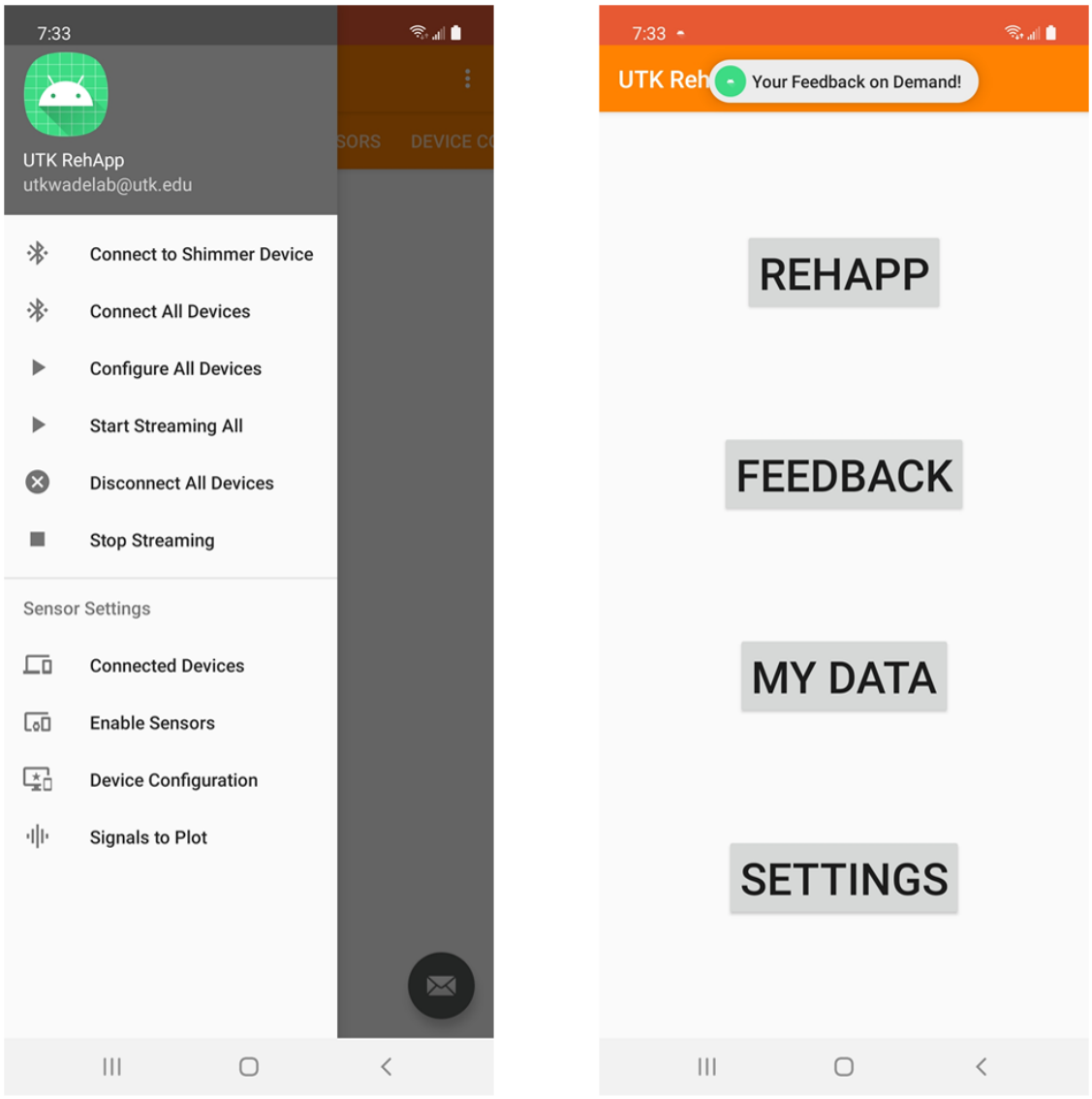


Figure 7.1: Window for connecting and configuring devices and starting streaming (a) and the homepage (b) with options to **REHAPP**) return to screen (a); **FEEDBACK**) request feedback as shown at the top of the screen; **MY DATA**) see the screen shown in **Figure 7.2(a)**; or **SETTING**) adjust settings.

of that classification were shown in the application, and that screen was captured and is displayed in **Figure 7.2**.

During this trial period, the right limb was designated as being "more affected" or non-dominant even though the participant is right hand dominant. This was done to test the rigor of the logic, wherein more results are available when the more affected limb is used proportionally more than the less affected limb. Of the $\tilde{5}000$ windows of data classified, 43% were classified as post-stroke (or compensatory) bimanual symmetric tasks. This result was followed by post-stroke unimanual tasks performed with the more affected limb (30%) and post-stroke bimanual asymmetric tasks performed with the less affected limb (16%) tasks. Only 2% of the task performances were classified as control (or healthy).

The participant's activities were recorded via camera throughout the trial period. The participant spent approximately 17% of the four hour period (41 minutes) folding laundry, approximately 57% of the trial baking cookies (137 minutes), 11% of the trial placing puzzle pieces with the "more affected" (dominant) limb (26 minutes), and 15% of the trial placing puzzle pieces with the "less affected" (non-dominant) limb (36 minutes). Considering the times captured on video to be "ground truth," **Table 7.2** shows the actual and predicted results of the analysis.

The participant also received feedback eleven times during the four hour trial period, and the interval and feedback given are shown in **Table 7.3**. The feedback was delivered based on the logic demonstrated in **Table 7.1**. Although the trial only took place over the course of four hours, the participant did not report any fatigue with the amount of feedback provided.

7.4 Discussion

This study was a proof of concept of a system that could deliver feedback to a participant regarding his task performance with the intent of promoting healthy use of the affected limb. The data were successfully packaged and sent to the server to be classified, and feedback was pushed to the participant based on logic supplied. However, there were many instances where difficulties arose or inaccuracies occurred.

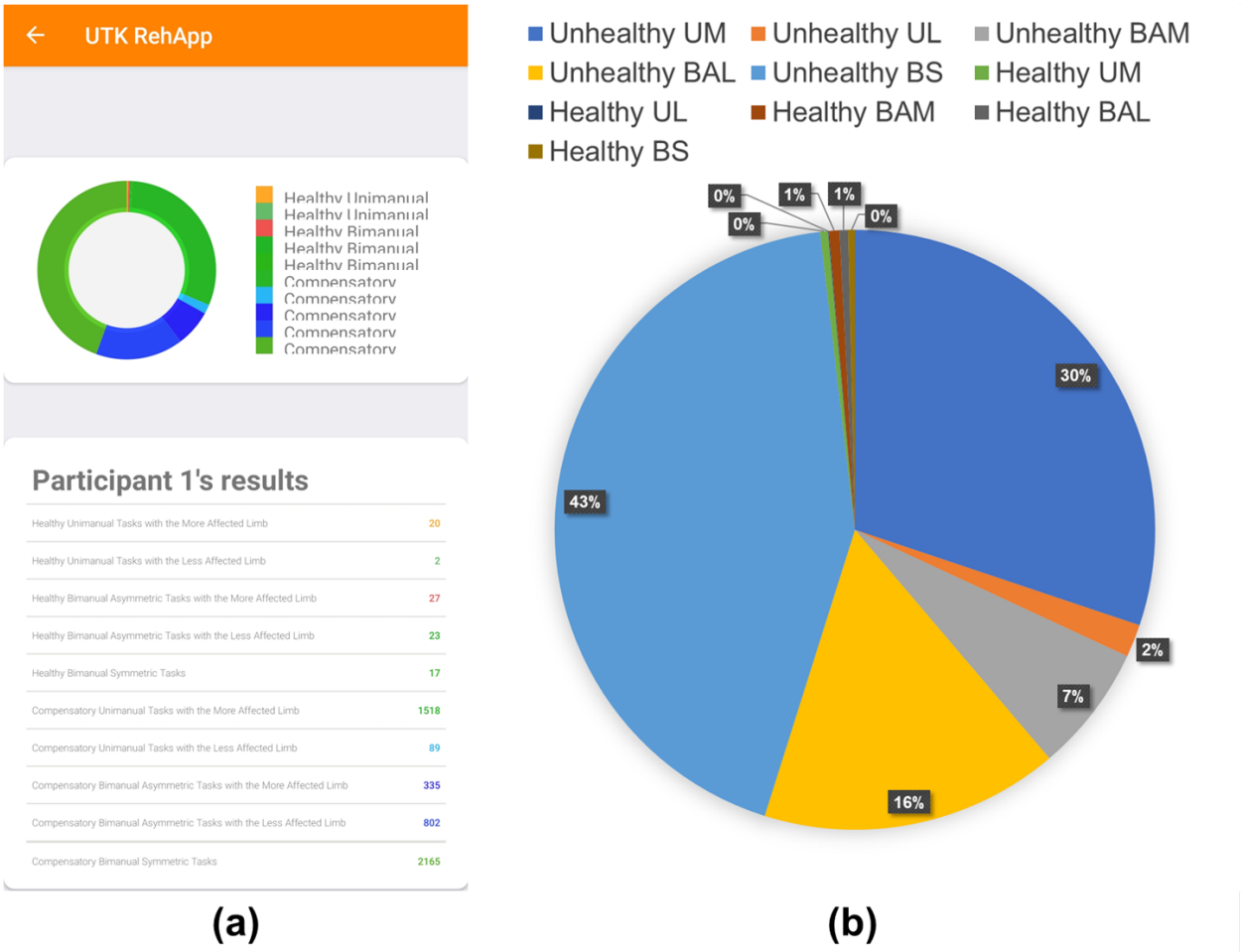


Figure 7.2: Results as shown in the application **(a)** and reorganized for clarity in the format of a pie chart **(b)**.

Table 7.2: Task performance divided by video and compared to predicted results

		Predicted			
		UM	**UL**	**BA**	**BS**
Actual	**UM**	101 (2.12%)	27 (0.57%)	138 (2.9%)	258 (5.42%)
	UL	277 (5.82%)	7 (0.15%)	183 (3.84%)	244 (5.13%)
	BA	835 (17.54%)	42 (0.88%)	632 (13.28%)	1221 (25.65%)
	BS	241 (5.06%)	10 (0.21%)	191 (4.01%)	353 (7.42%)

Table 7.3: Feedback delivered during trial

Interval	Feedback
1	During one-handed activity in the last time period, you used your affected hand more than your unaffected hand.
2	You demonstrated healthier activity than the previous period.
3	During two-handed activity, you used your affected hand more often this time period than the previous period.
4	You demonstrated healthier one-handed activity than the previous period.
5	You demonstrated healthier two-handed activity than the previous period.
6	Keep working toward healthy use of the affected limb!
7	During two-handed activity in this time period, your affected hand use was better than the previous period.
8	Your two-handed activity in this time period was better than the previous period.
9	During two-handed activity in the last time period, you used your affected hand more than your unaffected hand.
10	Keep working toward healthy use of the affected limb!
11	During one-handed activity in this time period, your affected hand use was better than the previous period.

In the initial development of the application, errors were discovered in the Android/Java API supplied by Shimmer. While the error was eventually addressed by the company, it significantly delayed the application development. Another delay came from the sudden expiration of one of the Shimmer sensor being used in development. Once this problem was solved, the application was significantly delayed in deployment, and we could not test the system on the target population of persons post-stroke. Then, even when the application was completed, it was difficult to manage connecting all five necessary sensors to the cellphone via Bluetooth, and when all the sensors were connected, the phone's battery quickly depleted over the course of the trial. Plugging in the phone resolved the power issue, but the phone also needed to remain connected to WiFi in order to send the packaged data to the server. This meant that the phone was tethered, either by power cord or WiFi signal, and the participant was tethered to the phone by the limits of the Bluetooth signal between the phone and sensors. This implies that unless changes are made, the system is not fully ready to be used in everyday activities of daily living.

Regardless of the limitations of the hardware, the classifier was also not producing desired results. The participant for this trial was a healthy control; however, 98% of task performances were classified as being post-stroke, or compensatory tasks. As mentioned in 7.2.2, the participant wore the sensors while performing ADLs, such as folding laundry, baking cookies, and building a puzzle. Often during the bimanual asymmetric task performance of baking cookies, the participant would perform tasks that would be considered unimanual of bimanual symmetric. Considering that for this trial, the "more affected" limb was designated as being the participant's dominant limb, the majority of unimanual tasks were UM. This would explain why the most prominent tasks are the bimanual symmetric tasks followed by the unimanual with the more affected limb. It is likely that, even though SMOTE combined with ENN was used in **Chapter 6** to prevent specific tasks from being overrepresented and remove and overlapping data, the post-stroke performance of each task was overrepresented. This lead to task performance being classified as compensatory, even though it was performed by a healthy control.

The feedback that was provided to the participant followed the logic exactly as it was designed. Even though this study was unable to vary the frequency with which feedback

was delivered to determine the optimal frequency for promoting healthy use of the affected limb, the participant did report that the feedback being delivered every 15 minutes did not result in intervention fatigue. In the future, longer trials will have to be performed with the target population to determine whether providing feedback at a higher frequency for a longer amount of time will result in frustration and fatigue with the system.

7.5 Conclusion

The purpose of this exploratory study was to show that an application was capable of measuring human motion and providing tailored feedback with the goal of reducing motor compensation. In this section, we showed that we were able to develop and application that was capable of streaming real-time data, packaging that data to send to a server to be classified, and delivering feedback based on the results of that classification. Unfortunately, classification accuracy of the trial period was low, so further work needs to be done to improve the accuracy of the ESNN on a novel dataset. This could be done by better balancing the datasets, either synthetically or by collecting more data. The system would then need to be piloted on the target populations of persons post-stroke. In the next chapter, we will summarize this research and provide some insights into the nature of the data used. This will lead us to a discussion of why some machine learning methods worked while others did not.

Chapter 8

Conclusion

8.1 Summary

As mentioned in the **Introduction**, the purpose of this project was to develop tools capable of reversing compensation and help persons post-stroke regain function through recovery. In **Chapter 2**, we showed that persons post-stroke use compensatory behaviors to regain lost function and independence, and that long-term use of these compensatory behaviors can lead to a decreased quality of life and lifespan. We discussed some therapeutic interventions designed to reverse compensatory behavior, such as CIMT, and their associated limitations. We also discussed how compensation can be measured in a laboratory setting for prototyped tasks. However, we discovered that such methods were not reasonable for application in the real world, or ambient setting, where compensation typically develops, so we expressed a need for minimally intrusive methods for remotely measuring compensation.

In **Chapter 3**, we evaluated the scope of the state-of-the-art literature on compensation to see how compensation is quantitatively described. In our review, we determined that, while researchers often describe the type of compensation observed in a trial, a quantitative definition of that compensation is rarely developed, let alone disseminated. We also learned that when quantitative descriptions of compensation are provided, they are provided in the context of a segment being used to compensate and a task being performed when compensation is measured. Finally, we learned that the nomenclature for compensatory

behaviors is varied, and we compiled a comprehensive list of named compensatory behaviors by segment.

Based on results from **Chapter 3**, our focus shifted to evaluating differences in task performance using wearable sensors. We began **Chapter 4** by describing the use of minimally intrusive sensors, triaxial IMUs, to collect data from participants post-stroke and age-matched controls while they performed tasks that were akin to tasks that would be performed in the ambient setting. These tasks were unimanual tasks performed with the more affected and less affected hand (UM and UL respectively), bimanual asymmetric tasks performed primarily with the more affected and less affected limb (BAM and BAL respectively), and bimanual symmetric tasks where both limbs were used equally and coordination between limbs was expected to be high (BS). We extracted representative features from these sensor data compared the features to determine whether there were significant differences between groups (post-stroke and control) and tasks, and these features revealed that there were statistically significant differences between both groups and tasks. Finally, we validated our findings by showing that specific extracted features correlate with clinical measures of impairment for participants post-stroke.

Because sensor data are sensitive to differences between group and task, in **Chapter 5**, we investigated methods for classifying the different tasks and groups. We first investigated supervised learning, which is commonly used for classification, including classifying human motion data, but can be limited to developing participant-specific models with smaller datasets. Using references from literature, we tested several supervised learning methods while iterating the number and type of features, the window size and overlap, and the methods of separating the data for validation and testing. We found that supervised learning methods had high success when the models were trained and tested with data from the same participant, but we were unable to create a generalized model capable of accurately classifying task and group. We then investigated other methods including unsupervised learning, statistical models, and different ways of preprocessing data to improve accuracy, but results never improved over the supervised learning methods. We found that the poor performance of these additional models was due to the features extracted from data, and high between-participant variability. Later in this chapter, we will discuss the nature of

our collected data and why each of the methods were unsuccessful in our attempts at classification.

Given the chaotic nature of our sensor, we investigated techniques appropriate for extracting features from chaotic time series signals. In **Chapter 6**, we found that recurrent neural networks (RNN) are capable of classifying complicated time series data. In our literature search, we found that the echo state neural network (ESNN) is an RNN that uses a loosely connected hidden layer (known as a reservoir) and works well with chaotic data. We tested a simple artificial neural network (ANN), known as a multi-layer perceptron (MLP), to establish a baseline accuracy for using neural networks to classify our data, and then we compared the results of the classification using the MLP to the ESNN. We found that the ESNN was capable of differentiating between tasks within groups(91.3% accuracy in control, 80.3% accuracy in post-stroke), but also being able to differentiate between groups and tasks(81.4% accuracy in combined control and post-stroke).

In **Chapter 7**, we describe the development of a system for remotely measuring compensation and providing feedback to encourage healthy use of the affected limb. To do this, we first developed an application that connects to the sensors and collects data for real-time classification using the ESNN from **Chapter 6**. We then wrote feedback and logic for the timing and delivery of said feedback to be delivered to the participant at specific intervals to provide insight into previous activity and encourage healthy use of the affected limb. The system was piloted with a single, non age-matched control to verify that the components were technically sound. The system demonstrated technical feasibility, but the classification accuracy on this single participant was low. The following chapter will also discuss why the tasks were inaccurately classified given the demonstrated success of the ESNN.

In this chapter, we characterize relevant aspects of our source data including 1) the chaotic nature of the data and 2) the small sample size. We will also discuss why methods used in **Chapter 5** were unsuccessful and what we learned from performing analyses using these methods. The nature of the ESNN, why it was successful in classifying the initially collected data, and why it was unsuccessful in classifying the data in pilot trial will be

discussed. Finally, we will discuss our conclusions and next steps for improved classification and quantification of compensation in the future.

8.2 Data Characterization

A chaotic system is defined as having aperiodic long-term behavior that exhibits sensitive dependence on initial conditions [33]. While this obviously applies to certain phenomena like the weather and the stock market, it can also apply to biological systems such as brain activity and motor function [147]. The best way to develop descriptive or predictive equations for such phenomena is to record time series data of one or a few representative variables and use deep data analysis to derive the equations that govern the system. While it is possible to account for and occasionally accurately predict trends in these systems, as demonstrated by the weather and the stock market, it is often impossible to accurately model this chaotic data.

Human motion is often considered to be nonlinear or chaotic and is modeled using such systems [23, 201, 236]. There are a variety of environmental and physiological inputs that factor into every motor function, and for any given person in any given environment, the same goal could be achieved differently. This is because, as described in the definition of chaotic systems, human motion is dependent on the initial conditions. From the activations which trigger motion in the brain, to bone and muscle structure, to learned behavior based on training or upbringing, the origins of human motion data are highly variable between persons and that leads to highly variable, or chaotic, data. Chen et al [41] summarized the problem, saying

> Human motion is dependent on many factors, such as the environment, culture, personal differences, and emotions. Different people will perform the same action differently, and even the same person will perform it differently at different times. Due to the large diversity of human body size, appearance, and shape, and the complexity of human actions, the task of automatically recognising actions is very challenging.

We observed this variability in **Section 5.4.2** when we extracted the representative features used to classify the given tasks and groups. The extracted AUC indicated that even though certain trends emerged for the given groups and tasks, the between participant variability was high enough to hide any significant group and task differences. As mentioned in that discussion, even though several different representative features were extracted from the data, there were no systemic differences that emerged between the groups and participants. It is possible that we were unable to find a significant trend in the data because of our limited sample size.

In addition to the variability that is demonstrated between participants is the dearth of human motion data that is available. The process for labelling human motion data can be laborious, so the amount and type of task performance available is limited. To add to this dilemma is a factor that we discussed in **Chapter 3**; when human motion data are collected by researchers, it is often done so by having a participant perform a prototyped task in a controlled setting, such as a lab. This means that even if a model were developed to represent some task performance, it might only be accurate under the same environmental conditions as the initial data collection. Ultimately, the combination of these two factors, the variability of the data and the small sample size, provided a complex problem that we attempted to solve using a variety of learning methods.

8.3 Performance of Machine Learning Approaches

As mentioned in **Section 5.3.1**, wearable sensors were used in [29] to differentiate between standing, walking, sitting and lying down using several different supervised learning methods. Supervised learning is commonly applied to human motion, including: characterizing gait [160] and walking conditions [296], recognizing specific movement [56, 153], and measuring activity [76] and function [198] is persons post-stroke. We used several supervised learning methods to classify the IMU data we had collected, iterating features selected, window size and overlap, as well as our cross-validation process. We found that we could create user specific models that were capable of classifying tasks, but the generalized models demonstrated poor performance. Low classification accuracy is not uncommon for cases like

this where a small dataset is used. There are generally believed to be two classification methods: uniform (where training data comes from all test subjects) and individual (where test data is from a separate subject or subjects than training data). As we saw, the uniform methods can lead to problems with model generalization while the individual method shrinks an already small training dataset [29]. Because we wanted to create generalized models which could be applied to any participant in any situation, we continued to use the individual activity classification method. This analysis revealed that adding more features did not continue to increase classification accuracy, that classification accuracy did not significantly change with varied window size given the limitations of the dataset, and that we needed to pursue approaches that allowed for the creation of more generalized models.

We then separated data based on the participants' Upper Extremity Fugl-Meyer (UEFM) scores and which limbs were affected. Because we observed high variability between participants, we believed that using methods to reduce that variability would increase classification accuracy. As discussed in **Section 5.4.2**, when separating participants post-stroke by their UEFM scores, classification accuracy did not significantly change. This is likely due to the fact that while the between participant variability decreased, so did the size of the training and testing sets. This is even more demonstrable in the separation by affected hand, which halved the training and testing sets for each model, since in this analysis the overall accuracy decreased. This is what led us to characterize the original data and led us to discover that the data were chaotic. Given the small sample size and the chaotic nature of the data, we began to look to neural networks for classification.

8.4 Neural Networks Performance

The multi-layer perceptron (MLP), has long been used for classifying human motion using data collected from sensors [163]. As mentioned in **Section 6.2.5**, the MLP has a relatively low computational cost and is frequently used in nonlinear time series prediction [155, 16, 242, 103]. Considering that our data is nonlinear, it is then unsurprising that a relatively high accuracy was achieved with the MLP when compared to the other machine learning methods. However, as we saw in **Section 6.3**, the best accuracy for the control participants

was approximately 20% higher than for the participants post-stroke. This implies that, while the MLP is capable of classifying nonlinear time series data, the more chaotic the data, like in the example of persons post-stroke, the less capable the MLP will be.

As mentioned in **Section 6.1**, the ESNN is a recurrent neural network that uses a loosely connected hidden layer (known as a reservoir) and works well with chaotic data [193, 271]. The ESNN has an input layer, a reservoir, and an output layer, though the central component of an ESNN is the reservoir in which neurons are connected randomly during the initialization of the ESNN [37, 239, 98]. After the ESNN is initialized, the input weights and recurrent weights in the reservoir are fixed, and only the output weights are trained. This, along with the way that the ESNN maps data for use in the reservoir, serves to speed up the process of training the model, it provides better classification when compared to other recurrent neural networks (RNNs), and it solves the vanishing gradient problem [91] that plagues other RNNs [47].

The ESNN has been used to classify EEG signal [115], to detect and index motion during tennis swings [12], and to drive a tracking controller for an exoskeleton [36]. When we were building the ESNN for our time series data, we had success in because the ESNN is capable of classifying chaotic time series data. We iterated the hyperparameters on our model, and the validation and testing sets were classified with a higher accuracy than we had seen in any of the previous methods. However, when we tested the ESNN on a novel dataset, the accuracy of the classification was low. This could be for a number of reasons. Firstly, it is possible that, despite our efforts in **Section 6.2.4** to balance the datasets, all of the post-stroke tasks were overrepresented leading to the majority of the tasks being classified as post-stroke or compensatory. This theory is supported by bimanual symmetric task performance taking a longer time for persons post-stroke and being highly classified in the proof of concept trial. Another possible explanation is that the model was overfitted on this data, which would explain why the classification accuracy was low on a novel dataset. A possible solution for this would be to collect more data and train a more generalized model, but as discussed in **Section 8.2**, it can be difficult to collect more human motion data. Finally, the differences in hardware between the initial data collection used to train the ESNN and the trial run

could have led to low classification accuracy. Not only were different sensors used, but it is possible that sensor location or orientation was off between the two data collection periods.

8.5 Future Work

The first consideration when determining next steps to improve performance is the balance of the source data. Whether this is done by collecting more data to add to the model or by synthetically balancing the data, it is the first step toward improving the classifier accuracy. Based on the success of the ESNN and the failures of other learning methods, it is reasonable to continue working with the ESNN as the method for classification. However, other avenues of time series classification might be pursued in attempting to achieve a higher and more generalized classification accuracy.

Another limitation of the current approach is the hardware. In order for this to be useful for real-time analysis, upgraded hardware would be needed for the system. Because one of the goals of this research was to use commercially available devices, the sensors used may not be the best for data collection and analysis. In the pilot trial, the participant had difficulty connecting all the sensors to the phone, keeping the phone charged and connected to WiFi, and dealing with constant crashes of the application. For this system to be evaluated in the future, the hardware would need to be upgraded to improve battery life, processing power, and connectivity between the sensors cellphone, and the system would need to be designed to work independent of any connection to the internet or the server so participants could move about their daily routine.

Finally, the whole system needs to be tested in a longer trial with the target population. The way that compensation is defined in this research is as being different from the control participant performance, but how that is implemented with persons post-stroke in a real-time trial needs to be investigated. A longer trial would also allow for the testing of different types of feedback and frequencies with which feedback could be delivered. Considering the goal of this project was to reduce compensation and promote function through recovery, it would be interesting to see how the whole system can work to do so.

Bibliography

[1] Afolabi, D., Guan, S.-U., Man, K. L., Wong, P. W., and Zhao, X. (2017). Hierarchical meta-learning in time series forecasting for improved interference-less machine learning. *Symmetry*, 9(11):283. 92

[2] Allen, L. N. and Christie, G. P. (2016a). The emergence of personalized health technology. *Journal of medical Internet research*, 18(5):e99. 2, 13

[3] Allen, L. N. and Christie, G. P. (2016b). The Emergence of Personalized Health Technology. *J. Med. Internet Res.*, 18(5):e99. 102

[4] Allen, M. T. (2018). A computer-based avatar task designed to assess behavioral inhibition extends to behavioral avoidance but not cognitive avoidance. *PeerJ*, 6:e5330. 7

[5] APDM (2020). Wearable Sensors - APDM Wearable Technologies. 67

[6] Aprile, I., Rabuffetti, M., Padua, L., Di Sipio, E., Simbolotti, C., and Ferrarin, M. (2014). Kinematic analysis of the upper limb motor strategies in stroke patients as a tool towards advanced neurorehabilitation strategies: a preliminary study. *Biomed Res Int*, 2014:636123. 5, 9, 17, 100

[7] Arksey, H. and O'Malley, L. (2005). Scoping studies: towards a methodological framework. *International journal of social research methodology*, 8(1):19–32. 19

[8] Arya, K. N. and Pandian, S. (2014). Interlimb neural coupling: implications for poststroke hemiparesis. *Ann Phys Rehabil Med*, 57(9-10):696–713. 40, 47, 100

[9] Asadi, H., Dowling, R., Yan, B., and Mitchell, P. (2014). Machine learning for outcome prediction of acute ischemic stroke post intra-arterial therapy. *PLoS One*, 9(2):e88225. 64

[10] Asadi, H., Kok, H. K., Looby, S., Brennan, P., O'Hare, A., and Thornton, J. (2016). Outcomes and Complications After Endovascular Treatment of Brain Arteriovenous Malformations: A Prognostication Attempt Using Artificial Intelligence. *World Neurosurg*, 96:562–569. 64

[11] Ávila, L. T., Chiviacowsky, S., Wulf, G., and Lewthwaite, R. (2012). Positive social-comparative feedback enhances motor learning in children. *Psychology of Sport and Exercise*, 13(6):849–853. 12, 104

[12] Bačić, B. (2016). Echo state network for 3d motion pattern indexing: A case study on tennis forehands. In Bräunl, T., McCane, B., Rivera, M., and Yu, X., editors, *Image and Video Technology*, pages 295–306, Cham. Springer International Publishing. 122

[13] Badami, R., VaezMousavi, M., Wulf, G., and Namazizadeh, M. (2012). Feedback about more accurate versus less accurate trials: Differential effects on self-confidence and activation. *Research quarterly for exercise and sport*, 83(2):196–203. 12, 104

[14] Bakhti, K., Laffont, I., Muthalib, M., Froger, J., and Mottet, D. (2018). Kinect-based assessment of proximal arm non-use after a stroke. *Journal of neuroengineering and rehabilitation*, 15(1):104. 8

[15] Bakhti, K., Mottet, D., Schweighofer, N., Froger, J., and Laffont, I. (2017). Proximal arm non-use when reaching after a stroke. *Neuroscience Letters*, 657:91–96. 8

[16] Bakker, R., Schouten, J. C., Giles, C. L., Takens, F., and Van Den Bleek, C. M. (2000). Learning chaotic attractors by neural networks. *Neural Computation*, 12(10):2355–2383. 94, 121

[17] Ballester, B. R., Maier, M., Mozo, R. M. S. S., Castañeda, V., Duff, A., and Verschure, P. F. (2016a). Counteracting learned non-use in chronic stroke patients with reinforcement-induced movement therapy. *Journal of neuroengineering and rehabilitation*, 13(1):74. 10

[18] Ballester, B. R., Maier, M., San Segundo Mozo, R. M., Castaneda, V., Duff, A., and M J Verschure, P. F. (2016b). Counteracting learned non-use in chronic stroke patients with reinforcement-induced movement therapy. *J Neuroeng Rehabil*, 13(1):74. 102

[19] Baniña, M. C., Mullick, A. A., McFadyen, B. J., and Levin, M. F. (2017). Upper limb obstacle avoidance behavior in individuals with stroke. *Neurorehabilitation and neural repair*, 31(2):133–146. 7, 8

[20] Barker, R. N. and Brauer, S. G. (2005). Upper limb recovery after stroke: the stroke survivors' perspective. *Disabil Rehabil*, 27(20):1213–1223. 8, 41

[21] Barker, R. N., Gill, T. J., and Brauer, S. G. (2007). Factors contributing to upper limb recovery after stroke: a survey of stroke survivors in Queensland Australia. *Disabil Rehabil*, 29(13):981–989. 8, 41

[22] Barth, J., Klaesner, J. W., and Lang, C. E. (2020). Relationships between accelerometry and general compensatory movements of the upper limb after stroke. *J Neuroeng Rehabil*, 17(1):138. 18

[23] Basharat, A. and Shah, M. (2009). Time series prediction by chaotic modeling of nonlinear dynamical systems. In *2009 IEEE 12th international conference on computer vision*, pages 1941–1948. IEEE. 87, 119

[24] Bates, D. W., Saria, S., Ohno-Machado, L., Shah, A., and Escobar, G. (2014). Big data in health care: using analytics to identify and manage high-risk and high-cost patients. *Health Aff (Millwood)*, 33(7):1123–1131. 64

[25] Batista, G. E. A. P. A., Prati, R. C., and Monard, M. C. (2004). A study of the behavior of several methods for balancing machine learning training data. *ACM SIGKDD Explorations Newsletter*, 6(1):20–29. 94

[26] Başţanlar, Y. and Ozuysal, M. (2014). Introduction to machine learning. *Methods Mol Biol*, 1107:105–128. 65

[27] Belfatto, A., Scano, A., Chiavenna, A., Mastropietro, A., Mrakic-Sposta, S., Pittaccio, S., Molinari Tosatti, L., Molteni, F., and Rizzo, G. (2018). A multiparameter approach to evaluate post-stroke patients: an application on robotic rehabilitation. *Applied Sciences*, 8(11):2248. 8

[28] Benjamin, E. J., Blaha, M. J., Chiuve, S. E., Cushman, M., Das, S. R., Deo, R., De Ferranti, S. D., Floyd, J., Fornage, M., Gillespie, C., et al. (2017). Heart disease and stroke statistics—2017 update. 4

[29] Bersch, S. D., Azzi, D., Khusainov, R., Achumba, I. E., and Ries, J. (2014). Sensor data acquisition and processing parameters for human activity classification. *Sensors (Basel)*, 14(3):4239–4270. 69, 70, 71, 92, 120, 121

[30] Bhosale, D. and Ade, R. (2014). Feature selection based classification using naive bayes, j48 and support vector machine. *International Journal of Computer Applications*, 99(16):14–18. 90

[31] Birkner, M. D., Kalantri, S., Solao, V., Badam, P., Joshi, R., Goel, A., Pai, M., and Hubbard, A. E. (2007). Creating diagnostic scores using data-adaptive regression: An application to prediction of 30-day mortality among stroke victims in a rural hospital in India. *Ther Clin Risk Manag*, 3(3):475–484. 64

[32] Bravata, D. M., Smith-Spangler, C., Sundaram, V., Gienger, A. L., Lin, N., Lewis, R., Stave, C. D., Olkin, I., and Sirard, J. R. (2007). Using pedometers to increase physical activity and improve health: a systematic review. *JAMA*, 298(19):2296–2304. 13

[33] Brock, W. (2001). Chaos theory. In Smelser, N. J. and Baltes, P. B., editors, *International Encyclopedia of the Social Behavioral Sciences*, pages 1643–1646. Pergamon, Oxford. 87, 119

[34] Brokaw, E. B., Holley, R. J., and Lum, P. S. (2013). Comparison of joint space and end point space robotic training modalities for rehabilitation of interjoint coordination in individuals with moderate to severe impairment from chronic stroke. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 21(5):787–795. 8

[35] Buma, F., Kwakkel, G., and Ramsey, N. (2013). Understanding upper limb recovery after stroke. *Restor. Neurol. Neurosci.*, 31(6):707–722. 5, 17, 100

[36] Cao, Y. and Huang, J. (2020). Neural-network-based nonlinear model predictive tracking control of a pneumatic muscle actuator-driven exoskeleton. *IEEE/CAA Journal of Automatica Sinica*, 7(6):1478–1488. 122

[37] Ceperic, V. and Baric, A. (2014). Reducing complexity of echo state networks with sparse linear regression algorithms. In *2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation*, pages 26–31. IEEE. 122

[38] Chartrand, G., Cheng, P. M., Vorontsov, E., Drozdzal, M., Turcotte, S., Pal, C. J., Kadoury, S., and Tang, A. (2017). Deep Learning: A Primer for Radiologists. *Radiographics*, 37(7):2113–2131. 88

[39] Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). Smote: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16:321–357. 94

[40] Chen, C. C. and Bode, R. K. (2010). Psychometric validation of the Manual Ability Measure-36 (MAM-36) in patients with neurologic and musculoskeletal disorders. *Arch Phys Med Rehabil*, 91(3):414–420. 44

[41] Chen, L., Wei, H., and Ferryman, J. (2013). A survey of human motion analysis using depth imagery. *Pattern Recognition Letters*, 34(15):1995–2006. Smart Approaches for Human Action Recognition. 119

[42] Chen, S., Wolf, S. L., Zhang, Q., Thompson, P. A., and Winstein, C. J. (2012). Minimal detectable change of the actual amount of use test and the motor activity log: the EXCITE Trial. *Neurorehabil Neural Repair*, 26(5):507–514. 18

[43] Chiviacowsky, S. and Harter, N. M. (2015). Perceptions of competence and motor learning: performance criterion resulting in low success experience degrades learning. *Brazilian Journal of Motor Behavior*, 9(1). 12, 104

[44] Chiviacowsky, S. and Wulf, G. (2007). Feedback after good trials enhances learning. *Research quarterly for exercise and sport*, 78(2):40–47. 12, 104

[45] Cho, S. I., Sun, S., Mun, J. H., Kim, C., Kim, S. Y., Cho, S., Youn, S. W., Kim, H. C., and Chung, J. H. (2020). Dermatologist-level classification of malignant lip diseases using a deep convolutional neural network. *Br J Dermatol*, 182(6):1388–1394. 64

[46] Chou, C.-Y., Ou, Y.-C., and Chiang, T.-R. (2015). Psychometric comparisons of four disease-specific health-related quality of life measures for stroke survivors. *Clinical Rehabilitation*, 29(8):816–829. 4

[47] Chouikhi, N., Ammar, B., and Alimi, A. M. (2018). Genesis of basic and multi-layer echo state network recurrent autoencoders for efficient data representations. *arXiv preprint arXiv:1804.08996*. 122

[48] Chow, J. W. and Stokic, D. S. (2014). Variability, frequency composition, and complexity of submaximal isometric knee extension force from subacute to chronic stroke. *Neuroscience*, 273:189–198. 49

[49] Chua, L. K., Wulf, G., and Lewthwaite, R. (2018). Onward and upward: Optimizing motor performance. *Hum Mov Sci*, 60:107–114. 12, 103

[50] Cirstea, M. C. and Levin, M. F. (2000). Compensatory strategies for reaching in stroke. *Brain*, 123 (Pt 5):940–953. 1, 5, 8, 61, 63

[51] Cirstea, M. C. and Levin, M. F. (2007). Improvement of arm movement patterns and endpoint control depends on type of feedback during practice in stroke survivors. *Neurorehabil Neural Repair*, 21(5):398–411. 11

[52] Clark, S. E. and Ste-Marie, D. M. (2007). The impact of self-as-a-model interventions on children's self-regulation of learning and swimming performance. *Journal of sports sciences*, 25(5):577–586. 12, 104

[53] Colquhoun, H. L., Levac, D., O'Brien, K. K., Straus, S., Tricco, A. C., Perrier, L., Kastner, M., and Moher, D. (2014). Scoping reviews: time for clarity in definition, methods, and reporting. *Journal of clinical epidemiology*, 67(12):1291–1294. 19

[54] Corbetta, D., Sirtori, V., Castellini, G., Moja, L., and Gatti, R. (2015). Constraint-induced movement therapy for upper extremities in people with stroke. *Cochrane Database Syst Rev*, (10):CD004433. 6

[55] Cortes, J. C., Goldsmith, J., Harran, M. D., Xu, J., Kim, N., Schambra, H. M., Luft, A. R., Celnik, P., Krakauer, J. W., and Kitago, T. (2017). A short and distinct time window for recovery of arm motor control early after stroke revealed with a global measure of trajectory kinematics. *Neurorehabilitation and neural repair*, 31(6):552–560. 8

[56] Cust, E. E., Sweeting, A. J., Ball, K., and Robertson, S. (2019). Machine and deep learning for sport-specific movement recognition: a systematic review of model development and performance. *Journal of Sports Sciences*, 37(5):568–600. PMID: 30307362. 120

[57] Debaere, F., Van Assche, D., Kiekens, C., Verschueren, S. M., and Swinnen, S. P. (2001). Coordination of upper and lower limb segments: deficits on the ipsilesional side after unilateral stroke. *Exp Brain Res*, 141(4):519–529. 2, 9, 41

[58] DeBow, S. B., Davies, M. L., Clarke, H. L., and Colbourne, F. (2003). Constraint-induced movement therapy and rehabilitation exercises lessen motor deficits and volume of brain injury after striatal hemorrhagic stroke in rats. *Stroke*, 34(4):1021–1026. 6, 101

[59] Del Din, S., Patel, S., Cobelli, C., and Bonato, P. (2011). Estimating Fugl-Meyer clinical scores in stroke survivors using wearable sensors. *Conf Proc IEEE Eng Med Biol Soc*, 2011:5839–5842. 42, 62

[60] Dettmers, C., Teske, U., Hamzei, F., Uswatte, G., Taub, E., and Weiller, C. (2005). Distributed form of constraint-induced movement therapy improves functional outcome and quality of life after stroke. *Arch Phys Med Rehabil*, 86(2):204–209. 6, 101

[61] Dipietro, L., Krebs, H. I., Volpe, B. T., Stein, J., Bever, C., Mernoff, S. T., Fasoli, S. E., and Hogan, N. (2012). Learning, not adaptation, characterizes stroke motor recovery: evidence from kinematic changes induced by robot-assisted therapy in trained and untrained task in the same workspace. *IEEE Trans Neural Syst Rehabil Eng*, 20(1):48–57. 18

[62] Dobkin, B. H., Plummer-D'Amato, P., Elashoff, R., and Lee, J. (2010). International randomized clinical trial, stroke inpatient rehabilitation with reinforcement of walking

speed (SIRROWS), improves outcomes. *Neurorehabil Neural Repair*, 24(3):235–242. 12, 103

[63] Doman, C. A., Waddell, K. J., Bailey, R. R., Moore, J. L., and Lang, C. E. (2016). Changes in Upper-Extremity Functional Capacity and Daily Performance During Outpatient Occupational Therapy for People With Stroke. *Am J Occup Ther*, 70(3):1–7003290040. 61, 62

[64] Dorsch, A. K., Thomas, S., Xu, X., Kaiser, W., and Dobkin, B. H. (2015a). Sirract: an international randomized clinical trial of activity feedback during inpatient stroke rehabilitation enabled by wireless sensing. *Neurorehabilitation and neural repair*, 29(5):407–415. 10

[65] Dorsch, A. K., Thomas, S., Xu, X., Kaiser, W., Dobkin, B. H., Emara, T., Edwards, D., Fonzetti, P., Maasch, J., Lee, S. G., Owolabi, M. O., Hamzat, T. K., LeBlanc, C. J., Morse, R., Swaminathan, N., Karatas, G. K., Boza, R., Brown, A. W., Miyai, I., Kawano, T., Chen, S. Y., Hanger, H., Zucconi, C., Mammi, S., Ghislanzoni, C., Juan, F., and Lang, C. E. (2015b). SIRRACT: An International Randomized Clinical Trial of Activity Feedback During Inpatient Stroke Rehabilitation Enabled by Wireless Sensing. *Neurorehabil Neural Repair*, 29(5):407–415. 102

[66] Epps, S., Doty, C., Knott, J., and Vokshi, A. (2013). Fitness wearables—many products, few customers. *Cambridge, MA: Forrester Reserach Inc.* 13

[67] Ericsson, A. (2015). Ericsson mobility report: On the pulse of the networked society. *Ericsson, Sweden, Tech. Rep. EAB-14*, 61078. 14

[68] Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., Cui, C., Corrado, G., Thrun, S., and Dean, J. (2019). A guide to deep learning in healthcare. *Nat Med*, 25(1):24–29. 65

[69] Figo, D., Diniz, P. C., Ferreira, D. R., and Cardoso, J. M. (2010). Preprocessing techniques for context recognition from accelerometer data. *Personal and Ubiquitous Computing*, 14(7):645–662. 68

[70] Fiore, L. D. and Lavori, P. W. (2016). Integrating Randomized Comparative Effectiveness Research with Patient Care. *N Engl J Med*, 374(22):2152–2158. 64

[71] Forouzanfar, M. H., Afshin, A., Alexander, L. T., Anderson, H. R., Bhutta, Z. A., Biryukov, S., Brauer, M., Burnett, R., Cercy, K., Charlson, F. J., et al. (2016). Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990–2015: a systematic analysis for the global burden of disease study 2015. *The Lancet*, 388(10053):1659–1724. 2, 13

[72] Fox, S. and Duggan, M. (2013). *Tracking for health*. Pew Research Center's Internet & American Life Project. 13

[73] Fradet, L., Lee, G., and Dounskaia, N. (2008a). Origins of submovements during pointing movements. *Acta Psychol (Amst)*, 129(1):91–100. 18

[74] Fradet, L., Lee, G., and Dounskaia, N. (2008b). Origins of submovements in movements of elderly adults. *J Neuroeng Rehabil*, 5:28. 18

[75] Fugl-Meyer, A. R., Jääskö, L., Leyman, I., Olsson, S., and Steglind, S. (1975). The post-stroke hemiplegic patient. 1. a method for evaluation of physical performance. *Scandinavian journal of rehabilitation medicine*, 7(1):13–31. 41, 44, 80

[76] Fulk, G. D. and Sazonov, E. (2011). Using sensors to measure activity in people with stroke. *Topics in Stroke Rehabilitation*, 18(6):746–757. PMID: 22436312. 120

[77] Gabrys, L., Sperzel, S., Bernhoerster, M., Banzer, W., and Vogt, L. (2017a). Real-time visual activity feedback for physical activity improvement in breast and colon cancer patients. *Research in Sports Medicine*, 25(1):1–10. 10

[78] Gabrys, L., Sperzel, S., Bernhoerster, M., Banzer, W., and Vogt, L. (2017b). Real-time visual activity feedback for physical activity improvement in breast and colon cancer patients. *Res Sports Med*, 25(1):1–10. 102

[79] Galloway, C. D., Valys, A. V., Petterson, F. L., Gundotra, V. P., Treiman, D. L., Albert, D. E., Dillon, J. J., Attia, Z. I., and Friedman, P. A. (2018). Non-invasive detection of

hyperkalemia with a smartphone electrocardiogram and artificial intelligence. *Journal of the American College of Cardiology*, 71(11S):A272–A272. 64

[80] Garry, M. I., van Steenis, R. E., and Summers, J. J. (2005). Interlimb coordination following stroke. *Hum Mov Sci*, 24(5-6):849–864. 40

[81] Gracia, D. (2013). Institute of medicine (iom). the learning healthcare system: Workshop summary. washington, dc: The national academies press, 2007. 64

[82] Haghgoo, H. A., Pazuki, E. S., Hosseini, A. S., and Rassafiani, M. (2013). Depression, activities of daily living and quality of life in patients with stroke. *J. Neurol. Sci.*, 328(1-2):87–91. 17, 100

[83] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. (2009). The weka data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1):10–18. 68, 70

[84] Hartman, J. M. (2007). Self-controlled use of a perceived physical assistance device during a balancing task. *Perceptual and motor skills*, 104(3):1005–1016. 12, 104

[85] Hawe, R. L. and Dewald, J. P. (2015). Development of a method to quantify inter-limb coupling in individuals with hemiparetic stroke. *Conf Proc IEEE Eng Med Biol Soc*, 2015:3476–3479. 47

[86] Hayward, K. S., Eng, J. J., Boyd, L. A., Lakhani, B., Bernhardt, J., and Lang, C. E. (2016). Exploring the role of accelerometers in the measurement of real world upper-limb use after stroke. *Brain Impairment*, 17(1). 18

[87] Heckman, B. W., Mathew, A. R., and Carpenter, M. J. (2015). Treatment Burden and Treatment Fatigue as Barriers to Health. *Curr Opin Psychol*, 5:31–36. 13, 104

[88] Held, J. P. O., Klaassen, B., Eenhoorn, A., van Beijnum, B. F., Buurke, J. H., Veltink, P. H., and Luft, A. R. (2018). Inertial Sensor Measurements of Upper-Limb Kinematics in Stroke Patients in Clinic and Home Environment. *Front Bioeng Biotechnol*, 6:27. 42

[89] Heo, J., Yoon, J. G., Park, H., Kim, Y. D., Nam, H. S., and Heo, J. H. (2019). Machine learning–based model for prediction of outcomes in acute stroke. *Stroke*, 50(5):1263–1265. 90

[90] Ho, K. C., Speier, W., El-Saden, S., Liebeskind, D. S., Saver, J. L., Bui, A. A., and Arnold, C. W. (2014). Predicting discharge mortality after acute ischemic stroke using balanced data. *AMIA Annu Symp Proc*, 2014:1787–1796. 64

[91] Hochreiter, S. (1998). The vanishing gradient problem during learning recurrent neural nets and problem solutions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 6(02):107–116. 122

[92] Hogan, N. and Sternad, D. (2009). Sensitivity of smoothness measures to movement duration, amplitude, and arrests. *J Mot Behav*, 41(6):529–534. 18

[93] Holmefur, M. M. and Krumlinde-Sundholm, L. (2016). Psychometric properties of a revised version of the Assisting Hand Assessment (Kids-AHA 5.0). *Dev Med Child Neurol*, 58(6):618–624. 41

[94] Hsu, W. C., Lau, K. H., Huang, R., Ghiloni, S., Le, H., Gilroy, S., Abrahamson, M., and Moore, J. (2016). Utilization of a Cloud-Based Diabetes Management Program for Insulin Initiation and Titration Enables Collaborative Decision Making Between Healthcare Providers and Patients. *Diabetes Technol. Ther.*, 18(2):59–67. 13

[95] Hu, T., Wu, F., and Zhang, X. (2007). Rainfall–runoff modeling using principal component analysis and neural network. *Hydrology Research*, 38(3):235–248. 90

[96] Huang, Y., Yang, Q., Chen, Y., and Song, R. (2017). Assessment of motor control during three-dimensional movements tracking with position-varying gravity compensation. *Frontiers in neuroscience*, 11:253. 9

[97] Hung, Y. C., Charles, J., and Gordon, A. M. (2004). Bimanual coordination during a goal-directed task in children with hemiplegic cerebral palsy. *Dev Med Child Neurol*, 46(11):746–753. 62

[98] Hunt, D., Parry, D., and Schliebs, S. (2014). Exploring the applicability of reservoir methods for classifying punctual sports activities using on-body sensors. In *ACSC*, pages 67–73. 122

[99] Hussain, I., Salvietti, G., Spagnoletti, G., Malvezzi, M., Cioncoloni, D., Rossi, S., and Prattichizzo, D. (2017). A soft supernumerary robotic finger and mobile arm support for grasping compensation and hemiparetic upper limb rehabilitation. *Robotics and Autonomous Systems*, 93:1–12. 9

[100] Hussain, I., Salvietti, G., Spagnoletti, G., and Prattichizzo, D. (2016). The soft-sixthfinger: a wearable emg controlled robotic extra-finger for grasp compensation in chronic stroke patients. *IEEE Robotics and Automation Letters*, 1(2):1000–1006. 9

[101] Hyman, S., Vogl, T., Blackwell, K., Barbour, G., Irvine, J., and Alkon, D. (1991). Classification of japanese kanji using principal component analysis as a preprocessor to an artificial neural network. In *IJCNN-91-Seattle International Joint Conference on Neural Networks*, volume 1, pages 233–238. IEEE. 90

[102] Ibáñez, J., Monge-Pereira, E., Molina-Rueda, F., Serrano, J. I., Del Castillo, M. D., Cuesta-Gómez, A., Carratalá-Tejada, M., Cano-de-la Cuerda, R., Alguacil-Diego, I. M., Miangolarra-Page, J. C., et al. (2017). Low latency estimation of motor intentions to assist reaching movements along multiple sessions in chronic stroke patients: a feasibility study. *Frontiers in neuroscience*, 11:126. 8

[103] Inoue, H., Fukunaga, Y., and Narihisa, H. (2001). Efficient hybrid neural network for chaotic time series prediction. In *International Conference on Artificial Neural Networks*, pages 712–718. Springer. 94, 121

[104] Janelle, C. M., Barba, D. A., Frehlich, S. G., Tennant, L. K., and Cauraugh, J. H. (1997). Maximizing performance feedback effectiveness through videotape replay and a self-controlled learning environment. *Research quarterly for exercise and sport*, 68(4):269–279. 12, 104

[105] Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., Wang, Y., Dong, Q., Shen, H., and Wang, Y. (2017a). Artificial intelligence in healthcare: past, present and future. *Stroke Vasc Neurol*, 2(4):230–243. 65

[106] Jiang, T., Gradus, J. L., and Rosellini, A. J. (2020). Supervised Machine Learning: A Brief Primer. *Behav Ther*, 51(5):675–687. 65

[107] Jiang, X., Merhi, L.-K., Xiao, Z. G., and Menon, C. (2017b). Exploration of force myography and surface electromyography in hand gesture classification. *Medical Engineering Physics*, 41:63–73. 90

[108] Johnson, N. B., Hayes, L. D., Brown, K., Hoo, E. C., and Ethier, K. A. (2014). Cdc national health report: leading causes of morbidity and mortality and associated behavioral risk and protective factors—united states, 2005–2013. 4

[109] Kaczmarczyk, K., Wit, A., Krawczyk, M., and Zaborski, J. (2009). Gait classification in post-stroke patients using artificial neural networks. *Gait Posture*, 30(2):207–210. 90

[110] Kamiali, A., Fister, I., Turkanovi, M., and Karakati, S. (2018). Sensors and Functionalities of Non-Invasive Wrist-Wearable Devices: A Review. *Sensors (Basel)*, 18(6). 14

[111] Kanai, M., Nozoe, M., Izawa, K. P., Takeuchi, Y., Kubo, H., Mase, K., and Shimada, S. (2017a). Promoting physical activity in hospitalized patients with mild ischemic stroke: a pilot study. *Topics in stroke rehabilitation*, 24(4):256–261. 10

[112] Kanai, M., Nozoe, M., Izawa, K. P., Takeuchi, Y., Kubo, H., Mase, K., and Shimada, S. (2017b). Promoting physical activity in hospitalized patients with mild ischemic stroke: a pilot study. *Top Stroke Rehabil*, 24(4):256–261. 102

[113] Kantak, S. S., Zahedi, N., and McGrath, R. L. (2016). Task-Dependent Bimanual Coordination After Stroke: Relationship With Sensorimotor Impairments. *Arch Phys Med Rehabil*, 97(5):798–806. 8, 42, 61, 62

[114] Kelso, J. A., Southard, D. L., and Goodman, D. (1979). On the coordination of two-handed movements. *J Exp Psychol Hum Percept Perform*, 5(2):229–238. 61

[115] Khan, Z. H., Hussain, N., and Tiwana, M. (2017). Classification of eeg signals for wrist and grip movements using echo state network. *Biomedical Research*, 28. 122

[116] Khusainov, R., Azzi, D., Achumba, I. E., and Bersch, S. D. (2013). Real-time human ambulation, activity, and physiological monitoring: taxonomy of issues, techniques, applications, challenges and limitations. *Sensors (Basel)*, 13(10):12852–12902. 67, 68

[117] King, G., Currie, M., and Petersen, P. (2014). Child and parent engagement in the mental health intervention process: A motivational framework. *Child and Adolescent Mental Health*, 19(1):2–8. 13, 104

[118] Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*. 95

[119] Kitago, T., Liang, J., Huang, V. S., Hayes, S., Simon, P., Tenteromano, L., Lazar, R. M., Marshall, R. S., Mazzoni, P., Lennihan, L., and Krakauer, J. W. (2013). Improvement after constraint-induced movement therapy: recovery of normal motor control or task-specific compensation? *Neurorehabil Neural Repair*, 27(2):99–109. 41, 100

[120] Klassen, T. D., Simpson, L. A., Lim, S. B., Louie, D. R., Parappilly, B., Sakakibara, B. M., Zbogar, D., and Eng, J. J. (2016a). "Stepping Up" Activity Poststroke: Ankle-Positioned Accelerometer Can Accurately Record Steps During Slow Walking. *Phys Ther*, 96(3):355–360. 102

[121] Klassen, T. D., Simpson, L. A., Lim, S. B., Louie, D. R., Parappilly, B., Sakakibara, B. M., Zbogar, D., and Eng, J. J. (2016b). "stepping up" activity poststroke: ankle-positioned accelerometer can accurately record steps during slow walking. *Physical therapy*, 96(3):355–360. 10

[122] Krabben, T., Prange, G. B., Molier, B. I., Stienen, A. H., Jannink, M. J., Buurke, J. H., and Rietman, J. S. (2012). Influence of gravity compensation training on synergistic

movement patterns of the upper extremity after stroke, a pilot study. *Journal of neuroengineering and rehabilitation*, 9(1):1–12. 9

[123] Krakauer, J. W. (2006). Motor learning: its relevance to stroke recovery and neurorehabilitation. *Curr. Opin. Neurol.*, 19(1):84–90. 1, 5

[124] Krumlinde-Sundholm, L., Holmefur, M., Kottorp, A., and Eliasson, A. C. (2007). The Assisting Hand Assessment: current evidence of validity, reliability, and responsiveness to change. *Dev Med Child Neurol*, 49(4):259–264. 41

[125] Krumlinde-Sundholm, L., Lindkvist, B., Plantin, J., and Hoare, B. (2019). Development of the assisting hand assessment for adults following stroke: a Rasch-built bimanual performance measure. *Disabil Rehabil*, 41(4):472–480. 41, 44

[126] Kumar, S., Nilsen, W. J., Abernethy, A., Atienza, A., Patrick, K., Pavel, M., Riley, W. T., Shar, A., Spring, B., Spruijt-Metz, D., et al. (2013). Mobile health technology evaluation: the mhealth evidence workshop. *American journal of preventive medicine*, 45(2):228–236. 12

[127] Kwakkel, G., Veerbeek, J. M., van Wegen, E. E., and Wolf, S. L. (2015). Constraint-induced movement therapy after stroke. *Lancet Neurol*, 14(2):224–234. 1, 6, 7, 101

[128] Köhler, S., Carmody, L., Vasilevsky, N., Jacobsen, J. O., Danis, D., Gourdine, J.-P., Gargano, M., Harris, N. L., Matentzoglu, N., McMurry, J. A., Osumi-Sutherland, D., Cipriani, V., Balhoff, J. P., Conlin, T., Blau, H., Baynam, G., Palmer, R., Gratian, D., Dawkins, H., Segal, M., Jansen, A. C., Muaz, A., Chang, W. H., Bergerson, J., Lauledeerkind, S. J., Yüksel, Z., Beltran, S., Freeman, A. F., Sergouniotis, P. I., Durkin, D., Storm, A. L., Hanauer, M., Brudno, M., Bello, S. M., Sincan, M., Rageth, K., Wheeler, M. T., Oegema, R., Lourghi, H., Della Rocca, M. G., Thompson, R., Castellanos, F., Priest, J., Cunningham-Rundles, C., Hegde, A., Lovering, R. C., Hajek, C., Olry, A., Notarangelo, L., Similuk, M., Zhang, X. A., Gómez-Andrés, D., Lochmüller, H., Dollfus, H., Rosenzweig, S., Marwaha, S., Rath, A., Sullivan, K., Smith, C., Milner, J. D., Leroux, D., Boerkoel, C. F., Klion, A., Carter, M. C., Groza, T., Smedley, D., Haendel, M. A.,

Mungall, C., and Robinson, P. N. (2018). Expansion of the Human Phenotype Ontology (HPO) knowledge base and resources. *Nucleic Acids Research*, 47(D1):D1018–D1027. 1, 4, 17, 100

[129] Lang, C. E., Bland, M. D., Bailey, R. R., Schaefer, S. Y., and Birkenmeier, R. L. (2013a). Assessment of upper extremity impairment, function, and activity after stroke: foundations for clinical decision making. *Journal of Hand Therapy*, 26(2):104–115. 5

[130] Lang, C. E., Bland, M. D., Bailey, R. R., Schaefer, S. Y., and Birkenmeier, R. L. (2013b). Assessment of upper extremity impairment, function, and activity after stroke: foundations for clinical decision making. *J Hand Ther*, 26(2):104–114. 17, 18, 100

[131] Lang, C. E., Wagner, J. M., Edwards, D. F., and Dromerick, A. W. (2007). Upper extremity use in people with hemiparesis in the first few weeks after stroke. *J Neurol Phys Ther*, 31(2):56–63. 18, 81, 102

[132] Lau, H.-Y., Tong, K.-Y., and Zhu, H. (2009). Support vector machine for classification of walking conditions of persons after stroke with dropped foot. *Human Movement Science*, 28(4):504–514. 90

[133] Lee, J. H., Ku, J., Cho, W., Hahn, W. Y., Kim, I. Y., Lee, S.-M., Kang, Y., Kim, D. Y., Yu, T., Wiederhold, B. K., et al. (2003). A virtual reality system for the assessment and rehabilitation of the activities of daily living. *CyberPsychology & Behavior*, 6(4):383–388. 7

[134] Lee, V. S. and Blanchfield, B. B. (2018). Disentangling Health Care Billing: For Patients' Physical and Financial Health. *JAMA*, 319(7):661–663. 64

[135] Levac, D., Colquhoun, H., and O'Brien, K. K. (2010). Scoping studies: advancing the methodology. *Implementation science*, 5(1):1–9. 19

[136] Levin, M. F. (1996). Interjoint coordination during pointing movements is disrupted in spastic hemiparesis. *Brain*, 119 (Pt 1):281–293. 8

[137] Levin, M. F., Kleim, J. A., and Wolf, S. L. (2009). What do motor "recovery" and "compensation" mean in patients following stroke? *Neurorehabil Neural Repair*, 23(4):313–319. 5, 17, 26, 41, 89, 100

[138] Levin, M. F., Liebermann, D. G., Parmet, Y., and Berman, S. (2016). Compensatory Versus Noncompensatory Shoulder Movements Used for Reaching in Stroke. *Neurorehabil Neural Repair*, 30(7):635–646. 41, 54, 100

[139] Lewthwaite, R. and Wulf, G. (2010). Social-comparative feedback affects motor skill learning. *Quarterly journal of experimental psychology*, 63(4):738–749. 12, 104

[140] Li, W.-J., Hsieh, C.-Y., Lin, L.-F., and Chu, W.-C. (2017). Hand gesture recognition for post-stroke rehabilitation using leap motion. *2017 International Conference on Applied System Innovation (ICASI)*. 90

[141] Li, X., Pan, X., Jiang, C., Wu, M., Liu, Y., Wang, F., Zheng, X., Yang, J., Sun, C., Zhu, Y., Zhou, J., Wang, S., Zhao, Z., and Zou, J. (2020). Predicting 6-Month Unfavorable Outcome of Acute Ischemic Stroke Using Machine Learning. *Front Neurol*, 11:539509. 64

[142] Liao, W., Wang, H., Zhang, J., Guo, C., Yao, J., and Jin, Y. (2019). An oil-immersed transformer fault diagnosis method based on data preprocessing and gradient boosting. In *2019 IEEE Power & Energy Society General Meeting (PESGM)*, pages 1–5. IEEE. 90

[143] Lima, R. C., Teixeira-Salmela, L., and Michaelsen, S. M. (2012). Effects of trunk restraint in addition to home-based modified constraint-induced movement therapy after stroke: a randomized controlled trial. *Int J Stroke*, 7(3):258–264. 6, 101

[144] Lin, J. H., Hsu, M. J., Sheu, C. F., Wu, T. S., Lin, R. T., Chen, C. H., and Hsieh, C. L. (2009). Psychometric comparisons of 4 measures for assessing upper-extremity function in people with stroke. *Phys Ther*, 89(8):840–850. 44, 80

[145] Lin, T.-H., Denomme, A., and Ranganathan, R. (2018). Learning alternative movement coordination patterns using reinforcement feedback. *Experimental brain research*, 236(5):1395–1407. 8

[146] Liu, W., McCombe Waller, S., Kepple, T. M., and Whitall, J. (2013). Compensatory arm reaching strategies after stroke: induced position analysis. *J Rehabil Res Dev*, 50(1):71–84. 5, 8, 17, 100

[147] Liu, Z. (2010). Chaotic Time Series Analysis. *Mathematical Problems in Engineering*, 2010:31 pages. 87, 119

[148] Livingston-Thomas, J. M., McGuire, E. P., Doucette, T. A., and Tasker, R. A. (2014a). Voluntary forced use of the impaired limb following stroke facilitates functional recovery in the rat. *Behavioural brain research*, 261:210–219. 10

[149] Livingston-Thomas, J. M., McGuire, E. P., Doucette, T. A., and Tasker, R. A. (2014b). Voluntary forced use of the impaired limb following stroke facilitates functional recovery in the rat. *Behav. Brain Res.*, 261:210–219. 102

[150] Lohse, K. R., Wulf, G., and Lewthwaite, R. (2012). Attentional focus affects movement efficiency. *Skill acquisition in sport: Research, theory & practice*, pages 40–58. 12, 104

[151] Long, X., Cleveland, W. L., and Yao, Y. L. (2005a). Effective automatic recognition of cultured cells in bright field images using fisher's linear discriminant preprocessing. *Image and vision computing*, 23(13):1203–1213. 90

[152] Long, X., Cleveland, W. L., and Yao, Y. L. (2005b). A new preprocessing approach for cell recognition. *IEEE Transactions on Information Technology in Biomedicine*, 9(3):407–412. 90

[153] Longstaff, B., Reddy, S., and Estrin, D. (2010). Improving activity classification for health applications on mobile devices using active and semi-supervised learning. In *2010 4th International Conference on Pervasive Computing Technologies for Healthcare*, pages 1–7. 120

[154] Love, A., Arnold, C. W., El-Saden, S., Liebeskind, D. S., Andrada, L., Saver, J., and Bui, A. A. (2013). Unifying acute stroke treatment guidelines for a Bayesian belief network. *Stud Health Technol Inform*, 192:1012. 64, 65

[155] Ma, Q., Zheng, Q., Peng, H., Zhong, T., and Xu, L. (2007). Chaotic time series prediction based on evolving recurrent neural networks. In *2007 International Conference on Machine Learning and Cybernetics*, volume 6, pages 3496–3500. 94, 121

[156] Mackenbach, J. P., Stirbu, I., Roskam, A. J., Schaap, M. M., Menvielle, G., Leinsalu, M., Kunst, A. E., van Oyen, H., Demarest, S., Rychtarikova, J., Dzurova, D., Andersen, O., Ekholm, O., Judge, K., Tekkel, M., Prattala, R., Martikainen, P., Desplanques, G., Jusot, F., Helmert, U., Kovacs, K., Marton, F., Layte, R., Costa, G., Vannoni, F., Villerusa, A., Kalediene, R., Klumbiene, J., Geurts, J. J., Dahl, E., Strand, B. H., Wojtyniak, B., Santana, P., Geckova, A. M., Artnik, B., Borrell, C., Esnaola, S., Regidor, E., Burstrom, B., Fritzell, J., Lundberg, O., Bopp, M., and Glickman, M. (2008). Socioeconomic inequalities in health in 22 European countries. *N. Engl. J. Med.*, 358(23):2468–2481. 2, 13

[157] Malcolm, M. P., Massie, C., and Thaut, M. (2009). Rhythmic auditory-motor entrainment improves hemiparetic arm kinematics during reaching movements: a pilot study. *Topics in stroke rehabilitation*, 16(1):69–79. 8

[158] Mandon, L., Boudarham, J., Robertson, J., Bensmail, D., Roche, N., and Roby-Brami, A. (2016). Faster reaching in chronic spastic stroke patients comes at the expense of arm-trunk coordination. *Neurorehabilitation and neural repair*, 30(3):209–220. 8

[159] Mannini, A., Trojaniello, D., Cereatti, A., and Sabatini, A. M. (2016a). A Machine Learning Framework for Gait Classification Using Inertial Sensors: Application to Elderly, Post-Stroke and Huntington's Disease Patients. *Sensors (Basel)*, 16(1). 64

[160] Mannini, A., Trojaniello, D., Cereatti, A., and Sabatini, A. M. (2016b). A machine learning framework for gait classification using inertial sensors: Application to elderly, post-stroke and huntington's disease patients. *Sensors*, 16(1). 120

[161] Mansfield, A., Wong, J. S., Bryce, J., Brunton, K., Inness, E. L., Knorr, S., Jones, S., Taati, B., and McIlroy, W. E. (2015a). Use of accelerometer-based feedback of walking activity for appraising progress with walking-related goals in inpatient stroke

rehabilitation: a randomized controlled trial. *Neurorehabilitation and neural repair*, 29(9):847–857. 10

[162] Mansfield, A., Wong, J. S., Bryce, J., Brunton, K., Inness, E. L., Knorr, S., Jones, S., Taati, B., and McIlroy, W. E. (2015b). Use of Accelerometer-Based Feedback of Walking Activity for Appraising Progress With Walking-Related Goals in Inpatient Stroke Rehabilitation: A Randomized Controlled Trial. *Neurorehabil Neural Repair*, 29(9):847–857. 102

[163] Mantyjarvi, J., Himberg, J., and Seppanen, T. (2001). Recognizing human motion with multiple acceleration sensors. In *2001 IEEE International Conference on Systems, Man and Cybernetics. e-Systems and e-Man for Cybernetics in Cyberspace (Cat.No.01CH37236)*, volume 2, pages 747–752 vol.2. 121

[164] Marchant, D. C. (2011). Attentional focusing instructions and force production. *Frontiers in psychology*, 1:210. 12, 104

[165] Marteniuk, R., MacKenzie, C., and Baba, D. (1984). Bimanual movement control: Information processing and interaction effects. *The Quarterly Journal of Experimental Psychology Section A*, 36(2):335–365. 61

[166] Martinez, C. A., Mintz, E., Ecsedy, A. E., and Fisher, B. E. (2017a). Constraining movement reveals motor capability in chronic stroke: an initial study. *Clinical rehabilitation*, 31(8):1126–1133. 10

[167] Martinez, C. A., Mintz, E., Ecsedy, A. E., and Fisher, B. E. (2017b). Constraining movement reveals motor capability in chronic stroke: an initial study. *Clin Rehabil*, 31(8):1126–1133. 102

[168] Massie, C. L. and Malcolm, M. P. (2012). Instructions emphasizing speed improves hemiparetic arm kinematics during reaching in stroke. *NeuroRehabilitation*, 30(4):341–350. 8

[169] Massie, C. L., Malcolm, M. P., Greene, D. P., and Browning, R. C. (2012). Kinematic motion analysis and muscle activation patterns of continuous reaching in survivors of stroke. *Journal of motor behavior*, 44(3):213–222. 8

[170] MATLAB & Simulink (2021). Butterworth filter design. 47

[171] Maudal, O. (1996). *Preprocessing data for neural network based classifiers: Rough sets vs Principal Component Analysis*. PhD thesis, Citeseer. 90

[172] McCallum, A., Nigam, K., and Ungar, L. H. (2000). Efficient clustering of high-dimensional data sets with application to reference matching. In *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 169–178. 90

[173] McCrea, P. H., Eng, J. J., and Hodgson, A. J. (2005). Saturated muscle activation contributes to compensatory reaching strategies after stroke. *Journal of neurophysiology*, 94(5):2999–3008. 8

[174] Merdler, T., Liebermann, D. G., Levin, M. F., and Berman, S. (2013). Arm-plane representation of shoulder compensation during pointing movements in patients with stroke. *J Electromyogr Kinesiol*, 23(4):938–947. 5, 17, 89, 100

[175] Michaelsen, S. M. and Levin, M. F. (2004). Short-term effects of practice with trunk restraint on reaching movements in patients with chronic stroke: a controlled trial. *Stroke*, 35(8):1914–1919. 1, 8

[176] Michaelsen, S. M., Luta, A., Roby-Brami, A., and Levin, M. F. (2001). Effect of trunk restraint on the recovery of reaching movements in hemiparetic patients. *Stroke*, 32(8):1875–1883. 8

[177] Michielsen, M. E., Selles, R. W., Stam, H. J., Ribbers, G. M., and Bussmann, J. B. (2012). Quantifying nonuse in chronic stroke patients: a study into paretic, nonparetic, and bimanual upper-limb use in daily life. *Arch Phys Med Rehabil*, 93(11):1975–1981. 2, 8, 9, 17, 41, 89, 100

[178] Miller, A., Duff, S. V., Quinn, L., Bishop, L., Youdan, G., Ruthrauff, H., and Wade, E. (2018). Development of Sensor-Based Measures of Upper Extremity Interlimb Coordination. *Conf Proc IEEE Eng Med Biol Soc*, 2018:3160–3164. 50, 67

[179] Miller, A. and Wade, E. (2021). Classifying Unimanual and Bimanual Upper Extremity Tasks in Individuals Post-Stroke. *Annu Int Conf IEEE Eng Med Biol Soc*, 2021. 103, 104

[180] Miller A, Quinn L, D. S. W. E. (2020). Comparison of Machine Learning approaches for Classifying Upper Extremity Tasks in Individuals Post-Stroke. *Annu Int Conf IEEE Eng Med Biol Soc.*, 2020:4330–4336. 89, 98, 103

[181] Mohanty, R., Sinha, A. M., Remsik, A. B., Dodd, K. C., Young, B. M., Jacobson, T., Mcmillan, M., Thoma, J., Advani, H., Nair, V. A., and et al. (2018). Machine learning classification to identify the stage of brain-computer interface therapy for stroke rehabilitation using functional connectivity. *Frontiers in Neuroscience*, 12. 90

[182] Mohri, M., Rostamizadeh, A., and Talwalkar, A. (2018). *Foundations of machine learning*. MIT press. 64, 65

[183] Mozaffarian, D., Benjamin, E. J., Go, A. S., Arnett, D. K., Blaha, M. J., Cushman, M., Das, S. R., De Ferranti, S., Després, J.-P., Fullerton, H. J., et al. (2016). Executive summary: heart disease and stroke statistics—2016 update: a report from the american heart association. *Circulation*, 133(4):447–454. 4

[184] Naghavi, N. and Wade, E. (2019). Prediction of freezing of gait in parkinson's disease using statistical inference and lower–limb acceleration data. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 27(5):947–955. 69

[185] Nahum-Shani, I., Smith, S. N., Spring, B. J., Collins, L. M., Witkiewitz, K., Tewari, A., and Murphy, S. A. (2018). Just-in-Time Adaptive Interventions (JITAIs) in Mobile Health: Key Components and Design Principles for Ongoing Health Behavior Support. *Ann Behav Med*, 52(6):446–462. 13, 103, 104

[186] National Academies of Sciences, E., Medicine, et al. (2015). *Improving diagnosis in health care*. National Academies Press. 64

[187] Nelson, L. A., Mulvaney, S. A., Gebretsadik, T., Ho, Y. X., Johnson, K. B., and Osborn, C. Y. (2016). Disparities in the use of a mHealth medication adherence promotion intervention for low-income adults with type 2 diabetes. *J Am Med Inform Assoc*, 23(1):12–18. 2, 14

[188] Nightingale, T. E., Walhin, J. P., Thompson, D., and Bilzon, J. L. (2015). Influence of accelerometer type and placement on physical activity energy expenditure prediction in manual wheelchair users. *PLoS ONE*, 10(5):e0126086. 67, 91

[189] Nudo, R. J. (2013). Recovery after brain injury: mechanisms and principles. *Front Hum Neurosci*, 7:887. 6, 101

[190] of Substance Abuse Unit, W. H. O. M. (2014). *Global status report on alcohol and health, 2014*. World Health Organization. 13

[191] Oh, S.-K., Kim, W.-D., and Pedrycz, W. (2016). Design of radial basis function neural network classifier realized with the aid of data preprocessing techniques: design and analysis. *International Journal of General Systems*, 45(4):434–454. 90, 166

[192] Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, 9(1):97–113. 44

[193] Ozturk MC, Xu D, P. J. (2007). Analysis and design of echo state networks. *Neural Comput.*, 19(1):111–38. 90, 122

[194] Page, S. J., Levine, P., and Hade, E. (2012). Psychometric properties and administration of the wrist/hand subscales of the Fugl-Meyer Assessment in minimally impaired upper extremity hemiparesis in stroke. *Arch Phys Med Rehabil*, 93(12):2373–2376. 44, 80

[195] Page, S. J., Sisto, S. A., and Levine, P. (2002). Modified constraint-induced therapy in chronic stroke. *Am J Phys Med Rehabil*, 81(11):870–875. 102

[196] Palmer, K., Chiviacowsky, S., and Wulf, G. (2016). Enhanced expectancies facilitate golf putting. *Psychology of Sport and Exercise*, 22:229–232. 12, 104

[197] Parnandi, A., Wade, E., and Mataric, M. (2010a). Motor function assessment using wearable inertial sensors. *Conf Proc IEEE Eng Med Biol Soc*, 2010:86–89. 42

[198] Parnandi, A., Wade, E., and Matarić, M. (2010b). Motor function assessment using wearable inertial sensors. In *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, pages 86–89. 120

[199] Parnandi, A.R., E. W. and Matarić, M. (2010). Motor function assessment using wearable inertial sensors. *IEEE EMBC*. 2, 9, 42, 89, 100

[200] Patel, M. S., Asch, D. A., and Volpp, K. G. (2015). Wearable devices as facilitators, not drivers, of health behavior change. *JAMA*, 313(5):459–460. 13

[201] Piorek, M. (2015). Chaotic properties of gait kinematic data. In *IFIP International Conference on Computer Information Systems and Industrial Management*, pages 111–119. Springer. 87, 119

[202] Portney, Leslie G., W. M. P. (2014). *Foundations of Clinical Research: Applications to Practice*. Pearson Education, Philadelphia, PA, USA, 3rd edition. 69

[203] Post, P. G., Fairbrother, J. T., Barros, J. A., and Kulpa, J. (2014). Self-controlled practice within a fixed time period facilitates the learning of a basketball set shot. *Journal of Motor Learning and Development*, 2(1):9–15. 12, 104

[204] Rabbi, M., Pfammatter, A., Zhang, M., Spring, B., and Choudhury, T. (2015). Automated personalized feedback for physical activity and dietary behavior change with mobile phones: a randomized controlled trial on adults. *JMIR Mhealth Uhealth*, 3(2):e42. 13

[205] Raghavan, P. (2015). Upper limb motor impairment after stroke. *Physical Medicine and Rehabilitation Clinics*, 26(4):599–610. 5

[206] Raghavan, P., Santello, M., Gordon, A. M., and Krakauer, J. W. (2010). Compensatory motor control after stroke: an alternative joint strategy for object-dependent shaping of hand posture. *J. Neurophysiol.*, 103(6):3034–3043. 8

[207] Rajkomar, A., Dean, J., and Kohane, I. (2019). Machine Learning in Medicine. *N Engl J Med*, 380(14):1347–1358. 64

[208] Ranganathan, R., Wang, R., Dong, B., and Biswas, S. (2017). Identifying compensatory movement patterns in the upper extremity using a wearable sensor system. *Physiological measurement*, 38(12):2222. 8

[209] Rehme, A. K., Volz, L. J., Feis, D. L., Bomilcar-Focke, I., Liebig, T., Eickhoff, S. B., Fink, G. R., and Grefkes, C. (2015). Identifying Neuroimaging Markers of Motor Disability in Acute Stroke by Machine Learning Techniques. *Cereb Cortex*, 25(9):3046–3056. 64

[210] Reiterer, V., Sauter, C., Klosch, G., Lalouschek, W., and Zeitlhofer, J. (2008). Actigraphy–a useful tool for motor activity monitoring in stroke patients. *Eur. Neurol.*, 60(6):285–291. 102

[211] Rhodes, H. (2014). Accessing and using data from wearable fitness devices. *J AHIMA*, 85(9):48–50. 14

[212] Roby-Brami, A., Feydy, A., Combeaud, M., Biryukova, E., Bussel, B., and Levin, M. (2003a). Motor compensation and recovery for reaching in stroke patients. *Acta neurologica scandinavica*, 107(5):369–381. 8

[213] Roby-Brami, A., Jacobs, S., Bennis, N., and Levin, M. F. (2003b). Hand orientation for grasping and arm joint rotation patterns in healthy subjects and hemiparetic stroke patients. *Brain research*, 969(1-2):217–229. 8

[214] Rohrer, B., Fasoli, S., Krebs, H. I., Hughes, R., Volpe, B., Frontera, W. R., Stein, J., and Hogan, N. (2002). Movement smoothness changes during stroke recovery. *J Neurosci*, 22(18):8297–8304. 18

[215] Roli, F. (2009). *Multiple Classifier Systems*, pages 981–986. Springer US, Boston, MA. 168

[216] Rose, D. K. and Winstein, C. J. (2013). Temporal coupling is more robust than spatial coupling: an investigation of interlimb coordination after stroke. *J Mot Behav*, 45(4):313–324. 9, 40, 41, 100

[217] Saemi, E., Porter, J. M., Ghotbi-Varzaneh, A., Zarghami, M., and Maleki, F. (2012). Knowledge of results after relatively good trials enhances self-efficacy and motor learning. *Psychology of Sport and Exercise*, 13(4):378–382. 12, 104

[218] Schmidt, R. and Lee, T. (1999). *Motor Control and Learning: A Behavioral Emphasis*. Human Kinetics. 10

[219] Schneeweiss, S. (2014). Learning from big health care data. *N Engl J Med*, 370(23):2161–2163. 64

[220] Schwartz, W. B., Patil, R. S., and Szolovits, P. (1987). Artificial intelligence in medicine. Where do we stand? *N Engl J Med*, 316(11):685–688. 64

[221] Seitz, R. J., Hildebold, T., and Simeria, K. (2011). Spontaneous arm movement activity assessed by accelerometry is a marker for early recovery after stroke. *J. Neurol.*, 258(3):457–463. 102

[222] Shechtman, O. and Sindhu, B. (2015). Clinical assessment recommendations. 44

[223] Shen, G., Jiaying, L., Peng, W., Jichuan, Z., and Ce, L. (2019). Bad data identification in power system based on optimized gsa algorithm. In *2019 Chinese Control And Decision Conference (CCDC)*, pages 5763–5767. IEEE. 90

[224] Shimmer (2021a). ConsensysPRO Software - Shimmer Wearable Sensor Technologies. 108

[225] Shimmer (2021b). Shimmer3 IMU Unit - Shimmer Wearable Sensor Technologies. 107

[226] Shimmer (2021c). Shimmer3 Java/Android API Shimmer Wearable Sensor Technologies. 108

[227] Sidiropoulos, A. N., Chen, S., Kaminski, T. R. M., and Gordon, A. M. (2019). Modulation of gait inter-limb coordination in children with unilateral spastic cerebral palsy after intensive upper extremity intervention. *Exp Brain Res*, 237(6):1409–1419. 62

[228] Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L., Lai, M., Bolton, A., Chen, Y., Lillicrap, T., Hui, F., Sifre, L., van den Driessche, G., Graepel, T., and Hassabis, D. (2017). Mastering the game of Go without human knowledge. *Nature*, 550(7676):354–359. 65

[229] Siripibal, N., Supratid, S., and Sudprasert, C. (2019). A comparative study of object recognition techniques: Softmax, linear and quadratic discriminant analysis based on convolutional neural network feature extraction. In *Proceedings of the 2019 International Conference on Management Science and Industrial Engineering*, pages 209–214. 90

[230] Sleimen-Malkoun, R., Temprado, J. J., Thefenne, L., and Berton, E. (2011). Bimanual training in stroke: How do coupling and symmetry-breaking matter? *BMC Neurol*, 11:11. 47

[231] Sousa, A. S., Silva, A., Santos, R., Sousa, F., and Tavares, J. M. (2013). Interlimb coordination during the stance phase of gait in subjects with stroke. *Arch Phys Med Rehabil*, 94(12):2515–2522. 9, 41

[232] Stein, J., Krebs, H. I., Frontera, W. R., Fasoli, S. E., Hughes, R., and Hogan, N. (2004). Comparison of two techniques of robot-aided upper limb exercise training after stroke. *Am J Phys Med Rehabil*, 83(9):720–728. 18

[233] Steinhubl, S. R., Kim, K. I., Ajayi, T., and Topol, E. J. (2018). Virtual care for improved global health. *Lancet*, 391(10119):419. 64

[234] Stewart, M. (2019). Predicting stock prices with echo state networks. 95

[235] Sukal-Moulton, T., Krosschell, K. J., Gaebler-Spira, D. J., and Dewald, J. P. (2014). Motor impairment factors related to brain injury timing in early hemiparesis. Part I: expression of upper-extremity weakness. *Neurorehabil Neural Repair*, 28(1):13–23. 41, 100

[236] Sun, Y. and Chen, J. (2021). Human movement recognition in dancesport video images based on chaotic system equations. *Advances in Mathematical Physics*, 2021. 87, 119

[237] Swinnen, S. P. and Carson, R. G. (2002). The control and learning of patterns of interlimb coordination: past and present issues in normal and disordered control. *Acta Psychol (Amst)*, 110(2-3):129–137. 40, 100

[238] Takeuchi, N. and Izumi, S. (2012). Maladaptive plasticity for motor recovery after stroke: mechanisms and approaches. *Neural Plast.*, 2012:359728. 6, 101

[239] Tang, F., Tiňo, P., and Chen, H. (2014). Learning the deterministically constructed echo state networks. In *2014 International Joint Conference on Neural Networks (IJCNN)*, pages 77–83. IEEE. 122

[240] Taub, E., Uswatte, G., Bowman, M. H., Mark, V. W., Delgado, A., Bryson, C., Morris, D., and Bishop-McKay, S. (2013). Constraint-induced movement therapy combined with conventional neurorehabilitation techniques in chronic stroke patients with plegic hands: a case series. *Arch Phys Med Rehabil*, 94(1):86–94. 102

[241] Taub, E., Uswatte, G., Mark, V. W., and Morris, D. M. (2006). The learned nonuse phenomenon: implications for rehabilitation. *Eura Medicophys*, 42(3):241–256. 6, 17, 41, 89, 100, 101

[242] Teo, K. K., Wang, L., and Lin, Z. (2001). Wavelet packet multi-layer perceptron for chaotic time series prediction: effects of weight initialization. In *International Conference on Computational Science*, pages 310–317. Springer. 94, 121

[243] Teskey, W. J., Elhabiby, M., and El-Sheimy, N. (2012). Inertial sensing to determine movement disorder motion present before and after treatment. *Sensors (Basel)*, 12(3):3512–3527. 9, 10, 42, 89, 100

[244] Thielman, G. (2010). Rehabilitation of reaching poststroke: a randomized pilot investigation of tactile versus auditory feedback for trunk control. *Journal of Neurologic Physical Therapy*, 34(3):138–144. 8

[245] Thielman, G. (2013). Insights into upper limb kinematics and trunk control one year after task-related training in chronic post-stroke individuals. *J Hand Ther*, 26(2):156–160. 8, 41, 54, 61, 100

[246] Thielman, G., Kaminski, T., and Gentile, A. (2008). Rehabilitation of reaching after stroke: comparing 2 training protocols utilizing trunk restraint. *Neurorehabilitation and neural repair*, 22(6):697–705. 8

[247] Thompson, A. K. (2012). Interlimb coordination during locomotion: finding available neural pathways and using them for gait recovery. *Clin Neurophysiol*, 123(4):635–637. 9, 41

[248] Timms, K. P., Rivera, D. E., Collins, L. M., and Piper, M. E. (2013). Control systems engineering for understanding and optimizing smoking cessation interventions. In *2013 American Control Conference*, pages 1964–1969. IEEE. 13

[249] Tirelli, T. and Pessani, D. (2011). Importance of feature selection in decision-tree and artificial-neural-network ecological applications. alburnus alburnus alborella: A practical example. *Ecological informatics*, 6(5):309–315. 90

[250] Tozlu, C., Edwards, D., Boes, A., Labar, D., Tsagaris, K. Z., Silverstein, J., Lane, H. P., Sabuncu, M. R., Liu, C., Kuceyeski, A., and et al. (2018). Predicting response to motor therapy in chronic stroke patients using machine learning. 90

[251] Trempe, M., Sabourin, M., and Proteau, L. (2012). Success modulates consolidation of a visuomotor adaptation task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38(1):52. 12, 104

[252] Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., Moher, D., Peters, M. D., Horsley, T., Weeks, L., et al. (2018). Prisma extension for scoping reviews (prisma-scr): checklist and explanation. *Annals of internal medicine*, 169(7):467–473. 20

[253] Trojaniello, D., Ravaschio, A., Hausdorff, J. M., and Cereatti, A. (2015). Comparative assessment of different methods for the estimation of gait temporal parameters using a single inertial sensor: application to elderly, post-stroke, Parkinson's disease and Huntington's disease subjects. *Gait Posture*, 42(3):310–316. 42

[254] Urbin, M., Bailey, R. R., and Lang, C. E. (2015a). Validity of body-worn sensor acceleration metrics to index upper extremity function in hemiparetic stroke. *Journal of neurologic physical therapy: JNPT*, 39(2):111. 10

[255] Urbin, M. A., Bailey, R. R., and Lang, C. E. (2015b). Validity of body-worn sensor acceleration metrics to index upper extremity function in hemiparetic stroke. *J Neurol Phys Ther*, 39(2):111–118. 102

[256] Uswatte, G., Foo, W. L., Olmstead, H., Lopez, K., Holand, A., and Simms, L. B. (2005). Ambulatory monitoring of arm movement using accelerometry: an objective measure of upper-extremity rehabilitation in persons with chronic stroke. *Arch Phys Med Rehabil*, 86(7):1498–1501. 102

[257] Uswatte, G., Giuliani, C., Winstein, C., Zeringue, A., Hobbs, L., and Wolf, S. L. (2006a). Validity of accelerometry for monitoring real-world arm activity in patients with subacute stroke: evidence from the extremity constraint-induced therapy evaluation trial. *Arch Phys Med Rehabil*, 87(10):1340–1345. 9, 10, 42, 89, 100, 102

[258] Uswatte, G. and Hobbs Qadri, L. (2009). A behavioral observation system for quantifying arm activity in daily life after stroke. *Rehabil Psychol*, 54(4):398–403. 18, 62

[259] Uswatte, G., Miltner, W. H., Foo, B., Varma, M., Moran, S., and Taub, E. (2000a). Objective measurement of functional upper-extremity movement using accelerometer recordings transformed with a threshold filter. *Stroke*, 31(3):662–667. 10

[260] Uswatte, G., Miltner, W. H., Foo, B., Varma, M., Moran, S., and Taub, E. (2000b). Objective measurement of functional upper-extremity movement using accelerometer recordings transformed with a threshold filter. *Stroke*, 31(3):662–667. 102

[261] Uswatte, G., Taub, E., Morris, D., Light, K., and Thompson, P. A. (2006b). The Motor Activity Log-28: assessing daily use of the hemiparetic arm after stroke. *Neurology*, 67(7):1189–1194. 18

[262] Utley, A. and Steenbergen, B. (2006). Discrete bimanual co-ordination in children and young adolescents with hemiparetic cerebral palsy: recent findings, implications and future research directions. *Pediatr Rehabil*, 9(2):127–136. 61, 62

[263] Valdés, B. A., Glegg, S. M., and Van der Loos, H. M. (2017). Trunk compensation during bimanual reaching at different heights by healthy and hemiparetic adults. *Journal of motor behavior*, 49(5):580–592. 8

[264] Van Gils, A., Meyer, S., Van Dijk, M., Thijs, L., Michielsen, M., Lafosse, C., Truyens, V., Oostra, K., Peeters, A., Thijs, V., Feys, H., Krumlinde-Sundholm, L., Kos, D., and Verheyden, G. (2018). The Adult Assisting Hand Assessment Stroke: Psychometric Properties of an Observation-Based Bimanual Upper Limb Performance Measurement. *Arch Phys Med Rehabil*, 99(12):2513–2522. 47

[265] van Kordelaar, J., van Wegen, E. E., and Kwakkel, G. (2012). Unraveling the interaction between pathological upper limb synergies and compensatory trunk movements during reach-to-grasp after stroke: a cross-sectional study. *Exp Brain Res*, 221(3):251–262. 8, 17, 89, 100

[266] Villar, J. R., González, S., Sedano, J., Chira, C., and Trejo-Gabriel-Galan, J. M. (2015). Improving human activity recognition and its application in early stroke diagnosis. *Int J Neural Syst*, 25(4):1450036. 64

[267] Virani SS, Alonso A, A. H. et al. (2021). American Heart Association Council on Epidemiology and Prevention Statistics Committee and Stroke Statistics Subcommittee. Heart Disease and Stroke Statistics-2021 Update: A Report From the American Heart Association. *Circulation*, 143(8):e254–e743. 1, 17, 89

[268] Wade, E., Chen, C., and Winstein, C. J. (2013). Spectral analyses of wrist motion in individuals poststroke: The development of a performance measure with promise for unsupervised settings. *N Neural Repair*, page In press. 42, 47

[269] Wade, E., Chen, C., and Winstein, C. J. (2014). Spectral analyses of wrist motion in individuals poststroke: the development of a performance measure with promise for unsupervised settings. *Neurorehabil Neural Repair*, 28(2):169–178. 49

[270] Wade, E. R. and Fan, T.-W. J. (2013). Kinematic measures of upper-extremity performance in the home setting for an individual post-stroke: A case report. In *2013 6th International IEEE/EMBS Conference on Neural Engineering (NER)*, pages 1262–1265. IEEE. 18

[271] Wang, L., Wang, Z., and Liu, S. (2016). An effective multivariate time series classification approach using echo state network and adaptive differential evolution algorithm. *Expert Systems with Applications*, 43:237–249. 90, 122

[272] Wang, Q., Markopoulos, P., Yu, B., Chen, W., and Timmermans, A. (2017a). Interactive wearable systems for upper body rehabilitation: a systematic review. *Journal of neuroengineering and rehabilitation*, 14(1):1–21. 10

[273] Wang, Q., Markopoulos, P., Yu, B., Chen, W., and Timmermans, A. (2017b). Interactive wearable systems for upper body rehabilitation: a systematic review. *J Neuroeng Rehabil*, 14(1):20. 102

[274] Wang, Z., Yan, W., and Oates, T. (2017). Time series classification from scratch with deep neural networks: A strong baseline. In *2017 International Joint Conference on Neural Networks (IJCNN)*, pages 1578–1585. 68, 95

[275] Washabaugh, E. P., Kalyanaraman, T., Adamczyk, P. G., Claflin, E. S., and Krishnan, C. (2017). Validity and repeatability of inertial measurement units for measuring gait parameters. *Gait Posture*, 55:87–93. 67

[276] Weigend, A. S., Mangeas, M., and Srivastava, A. N. (1995). Nonlinear gated experts for time series: Discovering regimes and avoiding overfitting. *International journal of neural systems*, 6(04):373–399. 68, 95

[277] Whitford, M., Schearer, E., and Rowlett, M. (2018). Effects of in home high dose accelerometer-based feedback on perceived and actual use in participants chronic post-stroke. *Physiother Theory Pract*, pages 1–11. 14

[278] Wiesendanger, M., Kaluzny, P., Kazennikov, O., Palmeri, A., and Perrig, S. (1994). Temporal coordination in bimanual actions. *Can. J. Physiol. Pharmacol.*, 72(5):591–594. 61

[279] Winstein, C. J. (1991). Knowledge of results and motor learning–implications for physical therapy. *Phys Ther*, 71(2):140–149. 11

[280] Winstein, C. J., Stein, J., Arena, R., Bates, B., Cherney, L. R., Cramer, S. C., Deruyter, F., Eng, J. J., Fisher, B., Harvey, R. L., Lang, C. E., MacKay-Lyons, M., Ottenbacher, K. J., Pugh, S., Reeves, M. J., Richards, L. G., Stiers, W., and Zorowitz, R. D. (2016). Guidelines for Adult Stroke Rehabilitation and Recovery: A Guideline for Healthcare Professionals From the American Heart Association/American Stroke Association. *Stroke*, 47(6):e98–e169. 17, 100

[281] Wittmann, F., Held, J. P., Lambercy, O., Starkey, M. L., Curt, A., Höver, R., Gassert, R., Luft, A. R., and Gonzenbach, R. R. (2016). Self-directed arm therapy at home after stroke with a sensor-based virtual reality training system. *Journal of neuroengineering and rehabilitation*, 13(1):75. 8

[282] Wolf, S. L. (2007). Revisiting constraint-induced movement therapy: are we too smitten with the mitten? Is all nonuse "learned"? and other quandaries. *Phys Ther*, 87(9):1212–1223. 7, 101

[283] Wolf, S. L., Catlin, P. A., Ellis, M., Archer, A. L., Morgan, B., and Piacentino, A. (2001). Assessing Wolf motor function test as outcome measure for research in patients after stroke. *Stroke*, 32(7):1635–1639. 41

[284] Wolf, S. L., Winstein, C. J., Miller, J. P., Taub, E., Uswatte, G., Morris, D., Giuliani, C., Light, K. E., and Nichols-Larsen, D. (2006). Effect of constraint-induced movement

therapy on upper extremity function 3 to 9 months after stroke: the EXCITE randomized clinical trial. *JAMA*, 296(17):2095–2104. 6

[285] Wolf, S. L., Winstein, C. J., Miller, J. P., Thompson, P. A., Taub, E., Uswatte, G., Morris, D., Blanton, S., Nichols-Larsen, D., and Clark, P. C. (2008). Retention of upper limb function in stroke survivors who have received constraint-induced movement therapy: the EXCITE randomised trial. *Lancet Neurol*, 7(1):33–40. 6

[286] Woo, J.-h., Song, J.-Y., and Choi, Y.-J. (2019). Performance enhancement of deep neural network using feature selection and preprocessing for intrusion detection. In *2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, pages 415–417. IEEE. 90

[287] Wulf, G. (2013). Attentional focus and motor learning: a review of 15 years. *International Review of sport and Exercise psychology*, 6(1):77–104. 12, 104

[288] Wulf, G., Chiviacowsky, S., and Lewthwaite, R. (2012). Altering mindset can enhance motor learning in older adults. *Psychology and Aging*, 27(1):14. 12, 104

[289] Wulf, G. and Lewthwaite, R. (2016). Optimizing performance through intrinsic motivation and attention for learning: The optimal theory of motor learning. *Psychonomic bulletin & review*, 23(5):1382–1414. 12, 103

[290] Wulf, G., Lewthwaite, R., Cardozo, P., and Chiviacowsky, S. (2018). Triple play: Additive contributions of enhanced expectancies, autonomy support, and external attentional focus to motor learning. *Q J Exp Psychol (Hove)*, 71(4):824–831. 2, 12, 103

[291] Wulf, G., Raupach, M., and Pfeiffer, F. (2005). Self-controlled observational practice enhances learning. *Research quarterly for exercise and sport*, 76(1):107–111. 12, 104

[292] Xu, C., Li, S., Wang, K., Hou, Z., and Yu, N. (2017). Quantitative assessment of paretic limb dexterity and interlimb coordination during bilateral arm rehabilitation training. *IEEE Int Conf Rehabil Robot*, 2017:634–639. 2, 9, 41

[293] Yang, C.-l., Creath, R. A., Magder, L., Rogers, M. W., and McCombe Waller, S. (2019a). Impaired posture, movement preparation, and execution during both paretic and nonparetic reaching following stroke. *Journal of neurophysiology*, 121(4):1465–1477. 8

[294] Yang, C. L., Creath, R. A., Magder, L., Rogers, M. W., and McCombe Waller, S. (2019b). Impaired posture, movement preparation, and execution during both paretic and nonparetic reaching following stroke. *J Neurophysiol*, 121(4):1465–1477. 41, 54, 100

[295] Ye, H., Shen, H., Dong, Y., et al. (2017). Using evidence-based medicine through advanced data analytics to work toward a national standard for hospital-based acute ischemic stroke treatment. *Mainland China*. 65

[296] yin Lau, H., yu Tong, K., and Zhu, H. (2009). Support vector machine for classification of walking conditions of persons after stroke with dropped foot. *Human Movement Science*, 28(4):504–514. 120

[297] Yozbatiran, N., Der-Yeghiaian, L., and Cramer, S. C. (2008). A standardized approach to performing the action research arm test. *Neurorehabil Neural Repair*, 22(1):78–90. 41

[298] Yu, L., Xiong, D., Guo, L., and Wang, J. (2016). A remote quantitative Fugl-Meyer assessment framework for stroke patients based on wearable sensor networks. *Comput Methods Programs Biomed*, 128:100–110. 42

Appendices

A Data Extraction Form

Data Extraction Form

Title _____

Authors _____

Journal _____ **Year** _____ **Vol** _____ **Pages** _____

Study Type _____

Include ☐ **Exclude** ☐

Reason for exclusion: _____

METHODS/TRIAL QUALITY

Does the article/paper/study examine **stroke**, the effects of stroke, or stroke rehabilitation? If no, what **neurodegenerative disease/** disorder does it examine?		
Does the paper examine **upper extremity**?		
Does the article/paper/study refer to **compensation/compensatory behavior**? How is it **defined**? (Broad definitions of compensation can typically be found in the **Introduction/Background** **e.g.** "...whereas compensation refers to using alternative strategies to accomplish a task.")		
If any of the above questions answered no, the paper does not meet the inclusion criteria, and no further analysis is necessary.		
Is/are specific form(s) of compensation specified? If yes, what category does it/do they fall into (mark all that apply)? (Can be found in **Background/Introduction**, **Methods** or **Discussion**. e.g. "...thereby minimizing truncal compensation." which would describe motor compensation due to inappropriate engagement of the trunk.)	**Non-Use:** Choosing (consciously or subconsciously) not to use the affect limb	
	Behavioral Compensation: Avoiding tasks/locations that provoke use of the affected limb	
	Motor Compensation: Engaging/recruiting inappropriate body segments to perform a task	
	Other Compensation: Anything not listed	

If Motor Compensation:	
What segment is **inappropriately recruited**, or which segments are **inappropriately paired**? (Can be found in **Background/Introduction**, **Methods** or **Discussion**. e.g. "...in adaptive compensatory movements of the trunk..." which indicates inappropriate trunk involvement)	
Is the task that is being performed when compensation is observed defined? (Can be observed in **Methods**.)	
If Other Compensation:	
What is the compensation being defined?	

Was the defined compensation quantified? (Can be found in **Methods**, **Results** or **Discussion**. e.g. "...were used to quantify paretic and nonparetic UE amount of use...")	
What devices are being used to quantify compensation? (Can be found in **Methods**, **Results** or **Discussion**. e.g. "Participants chronic post-stroke wore bilateral **triaxial wrist accelerometers**...")	
What numerical quantities are being used to evaluation levels of compensation? (Can be found in **Methods**, **Results** or **Discussion**. e.g. "**Accelerometer data** were sampled data at 30 Hz ... to compute the number of **activity**) **counts**...")	
In what environment is this compensation being evaluated? (Can be found in **Methods**, **Results** or **Discussion**. e.g. "...feedback was administered at seven sessions in the **home setting**...")	

Participant • Age (Mean, range) • Other sample characteristics • Subject inclusion/exclusion criteria	
Method of subject selection	
Method of group assignment (randomization) Was allocation concealed?	
Study Design	
Blinding • Participants • Investigators • Outcome assessors	
Type of intervention • Intervention(s) • Control condition(s) • Duration and other protocol information	
Intention-to-treat analysis	
Outcome assessments	
Compliance	
Match of intervention and controls	
Baseline similarly between groups	

Results

	Experimental Group(s)	**Control Group(s)**
Observed n		
Excluded Subjects		
Lost to follow-up		
Lost to follow-up • Mean, standard-deviation • Proportion • Other effect size index		
Secondary Outcomes		
Statistical Analysis • Description of groups • Comparison of groups		

B Data Separation by Band

As mentioned in **Section 6.1**, data preprocessing techniques have been used in conjunction with neural networks to improve the results of classification analysis. Literature shows that some clustering and dimensionality reduction techniques can improve the accuracy of neural network classifiers [191]. In order to investigate this approach with upper extremity motion data, we evaluated a preprocessing technique to examine its effects on the accuracy of the MLP used to establish a baseline accuracy in **Chapter 6**.

B.1 Methods

In **Chapter 4**, the integral (or AUC) was correlated with clinical measures, and in the exploration of the raw data in **Chapter 5**, despite some of the grouping hiding significant difference, some distinctions between tasks and groups were empirically observed in the AUC analysis. To cluster the data, the normalized area under the curve (AUC) was calculated for each limb during task performance. The AUC for the dominant or less affected (LA) limb (represented by $(AUC_{LAW} + AUC_{LAU})/T(s)$, where $T(s)$ is the time required to complete the task in seconds) was plotted versus the area under the curve of the non-dominant or more affected (MA) limb (represented by $(AUC_{MAW} + AUC_{MAU})/T(s)$). Control participant data (**Figure 1(a)**) shows the tasks falling into five distinct regions (indicated by the numerically labelled spaces between the lines on **Figures 1(a) and 1(b)** and hereafter referred to as bands I-V) which do not necessarily correspond with task type. The first "band" (labelled I on **Figure 1(a)**) primarily contains the two unimanual tasks (reach for spoon or bottle) performed with the less affected limb. The second band (II) primarily contains the bimanual asymmetric task stir with less affected limb. The third band (III) contains both bimanual symmetric tasks (fold towel and don hat) and the two bimanual asymmetric tasks open bottle with either limb. The fourth band (IV) contains the bimanual asymmetric task stir with the more affected limb, and the fifth band (V) contains the two unimanual tasks performed with the more affected limb.

Since empirical analysis revealed that, for control participants, the data were separated into unique clusters by task type when plotting the area under the curve of the less affected

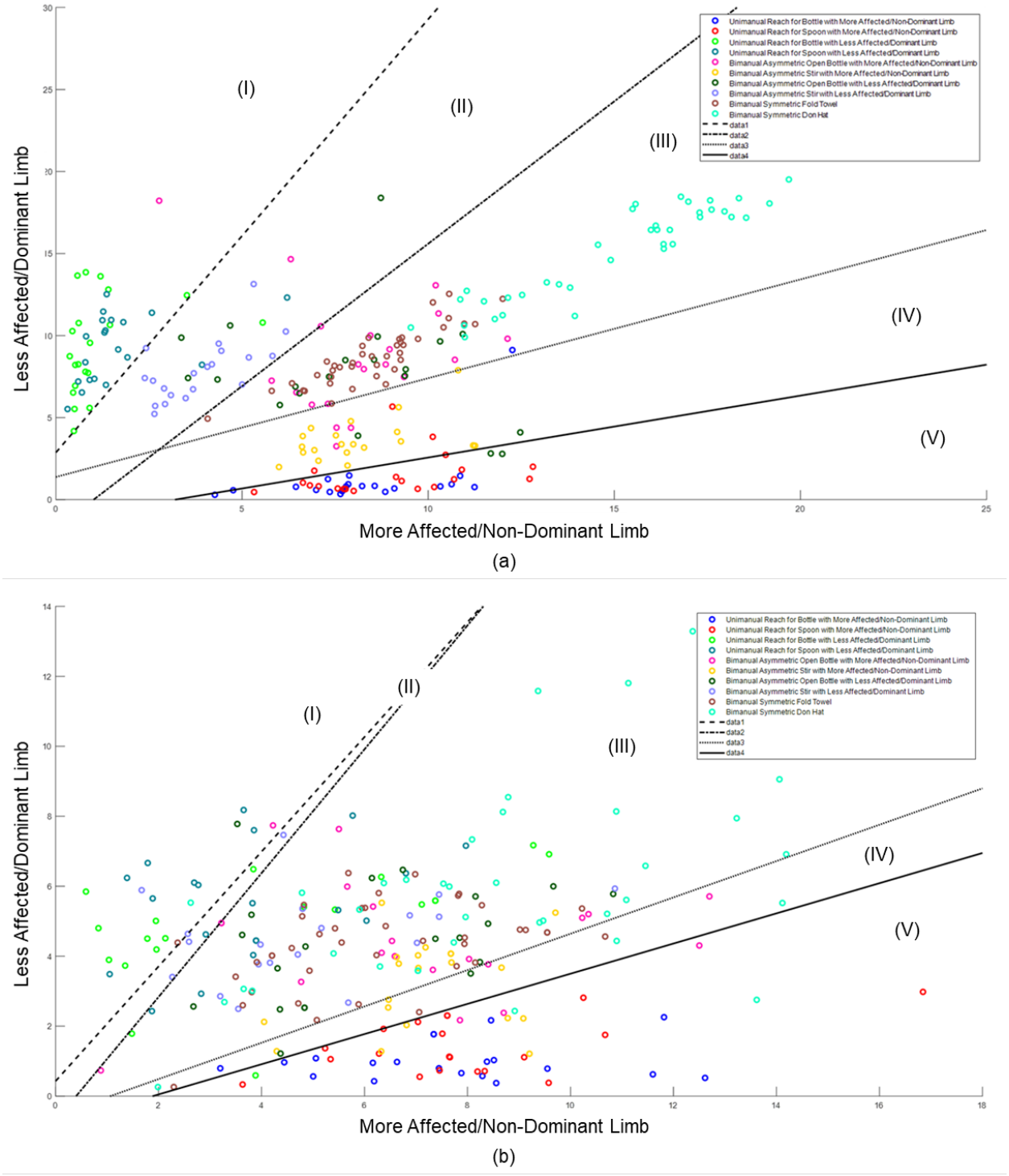


Figure 1: Area Under the Curve of the Less Affected Limb plotted versus the Area Under the Curve of the More Affected Limb for Control (a) and Post-Stroke (b) Groups.

or dominant limb versus the more affected or non-dominant limb, each task was assigned as "belonging" to the cluster, or band, where it primarily appeared. Based on that classification, the precision and recall for each band was evaluated for all control participants. The results of that analyses are shown in **Table 1**. The same analysis was done for participants post-stroke, manually changing the shape of the bands to optimize the precision, recall, and F1 score. The shape of the bands is shown in **Figure 1(b)**, and the final precision, recall, and F1 scores for each band are shown in **Table 2**.

Prior data analyses (**Chapter 5**) indicated that models for distinguishing participant group and task were not sufficiently accurate. Preprocessing the data by separating into bands allows for separate classifiers to be created for each band, reducing the computational requirements of each individual classifier. The process for separating and classifying the data includes: 1) assigning data to a band based on the results of the AUC analysis for the control participants (**I**: Both UL; **II**: BAL - Stir; **III**: Both BS and BAL & BAM - Unscrew; **IV**: BAM - Stir; and **V**: Both UM); 2) the dataset, now labelled by band instead of task performed, is used to train a new classifier to sort data into one of the five bands; 3) having verified the dataset has been correctly placed in a band, one of five secondary classifiers specific to the given band is used to classify the task. While this is not technically an ensemble method [215], such as the process used in decision trees, it allows for the most important features to be used and questions to be asked at each layer, increasing the likelihood of higher classification accuracy.

B.2 Results

When preprocessing the data by separating into empirically determined bands, there are multiple instances for accuracy loss as demonstrated in **Figure 2**. The first instance is the initial sorting into an appropriate band (**Figure 2(a)**); the second instance, which only applies when the data is not initially sorted into the appropriate band, is the second attempt at sorting the data into the correct band from the incorrect band (**Figure 2(b)**); and the third instance, sorting the data into the appropriate task using the classifier specific to each band (**Figure 2(c)**). The results of this analysis of the preprocessed data for control participants are shown in **Table 3**.

Table 1: Precision, recall, and F1 scores for preprossessing of control participants' data.

Control			
Band	**Precision**	**Recall**	**F1-Score**
Band 1	0.9737	0.925	0.9487
Band 2	0.6897	1	0.8163
Band 3	0.9907	0.8833	0.9339
Band 4	0.7143	1	0.8333
Band 5	0.9474	0.9	0.9231

Table 2: Precision, recall, and F1 scores for preprossessing of post-stroke participants' data.

Post-Stroke			
Band	**Precision**	**Recall**	**F1-Score**
Band 1	0.7619	0.4	0.5246
Band 2	0.6667	0.2	0.3077
Band 3	0.6824	0.8417	0.7537
Band 4	0.2667	0.2	0.2286
Band 5	0.78	0.975	0.8667

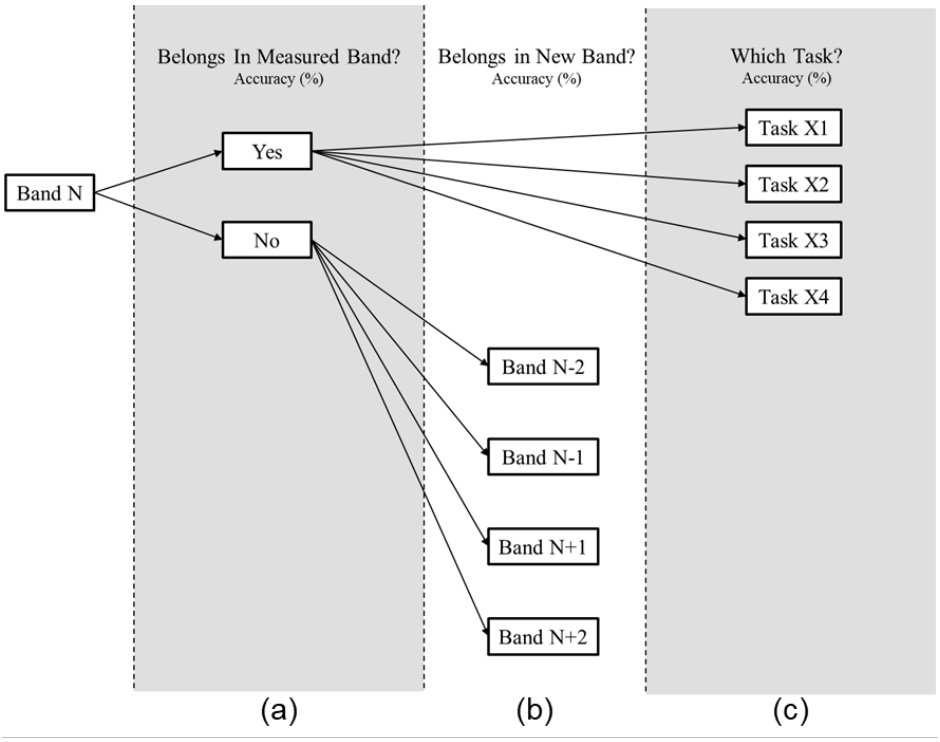


Figure 2: Results of preprocessing the data by separating into bands. A classifier is used to sort data into a band (a); if the initial classification was incorrect, the data is then sorted to the correct band (b); finally, each band has a specific classifier which classifies the tasks for that band (c).

Table 3: Loss, Accuracy, Validation Loss, and Validation Accuracy of multi-layer perceptron analysis of preprocessed control data.

Data	**Loss**	**Accuracy**	**Validation Loss**	**Validation Accuracy**
All Data	0.2584	0.9182	2.6455	**0.6006**
Band 1	0.5178	0.8828	6.1204	0.411
Band 2	0.1718	0.9564	1.8156	0.7172
Band 3	0.1788	0.9415	1.288	0.7451
Band 4	0.0896	0.977	2.4348	0.6107
Band 5	0.4417	0.8507	2.2561	0.6429
Bands Average	0.9324	0.975	2.1697	**0.6714**

As shown in the table, preprocessing the data by separating it into bands does not significantly improve the accuracy for control participants. Given that the F1-Scores for post-stroke participants were much lower than the control participants when the data were separated into bands and the computational effort required to analyze the data, no analysis was done for participants post-stroke.

B.3 Discussion

As others have shown, preprocessing data can be useful in improving the classifier accuracy. While the classifier accuracy did not significantly increase for our control participants, the accuracy of the classification were higher when the data was separated into bands. However, the computational requirements of the analysis and the insignificant improvement discouraged further testing of the preprocessing technique. Empirical analysis of **Figure 1(a)**, plotting the area under the curve for the dominant limb by the area under the curve of the non-dominant limb for control participants, demonstrated that combining all the asymmetric tasks may have been masking some significant differences and decreasing the accuracy of classification. However, analysis of **Figure 1(b)** representing participants post-stroke did not provide the granularity and insight into the data as represented by the difference in F1 scores between both groups of participants. Further efforts could be made to improve preprocessing of the data to increase validation accuracy, but the marked improvement of the ESNN in **Chapter 6** over previous work and the computational costs need to be taken into consideration.

Vita

Aaron Miller was born in Clarksville, TN on March 29th, 1994. In 1998, he moved to Crossville, TN where he attended Stone Elementary School until May 2008. In May 2012, he graduated 4th in his class from Cumberland County High School. In August of the same year, he began pursuing a Bachelor's Degree in Biomedical Engineering at the University of Tennessee, Knoxville.

During his time at the University of Tennessee, Aaron was a member of the Chancellor's Honors Program and served as a student representative on the Honors Council. He also worked as an Intern in the Higher Education Research Experiences (HERE) program at Oak Ridge National Lab (ORNL) for two years while pursuing his degree. In May of 2016, he graduated Cum Laude with a B.S. in Biomedical Engineering followed by a B.S. in Mechanical Engineering in December 2016.

In January of 2017, Aaron started as a graduate research assistant (GRA) in the Behavior, Robotics and Sensing Group at the University of Tennessee, Knoxville under the advisement of Dr. Eric Espinoza-Wade. As a GRA, Aaron worked on the NIH funded grant, "Quantifying and Reducing Motor Compensation After Stroke in Ambient Settings." He also worked as an advisor and instructor for the Governor's School for Science and Engineering (GSSE) and the High School Introduction to Engineering Systems (HITES) programs at UTK. He graduated with his Master of Science Degree in Biomedical Engineering in May 2018 and his Doctor of Philosophy Degree in Biomedical Engineering in December 2021.