# UC Irvine

## UC Irvine Electronic Theses and Dissertations

### Title

Novel Algorithms for Improving Wearable Feedback after Stroke

### Permalink

https://escholarship.org/uc/item/48c1122t

### Author

Okita, Shusuke

### Publication Date

2023

### Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, availalbe at https://creativecommons.org/licenses/by/4.0/

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA,
IRVINE

# Novel Algorithms for Improving Wearable Feedback after Stroke

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Mechanical and Aerospace Engineering

by

Shusuke Okita

Dissertation Committee:
Professor David J. Reinkensmeyer, Chair
Professor Terence Sanger
Professor Alexandra S. Voloshina
Professor An H. Do (COI oversight)

2023

© 2023 Shusuke Okita

# DEDICATION

To

my parents and brothers.

And to all the other people who have encouraged and inspired me along the way.

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# ACKNOWLEDGMENTS

Firstly, I would like to express my profound gratitude to my supervisor, Dr. David J. Reinkensmeyer, for his invaluable guidance, unwavering support, and constant encouragement throughout my PhD journey. I am deeply grateful for his steadfast belief in me and for inspiring me to strive for excellence. I'm honored to have been a part of BioRobotics Laboratory. During the program, there were numerous challenging moments, but your warm support, comprehensive feedback, and mentorship helped me persevere. From you, I have learned critical thinking, the art of formulating research questions, effective ways to answer questions, and how to organize and manage a research group. Your compassionate support, even in the face of adversity, has made me feel like you are my father in the US.

Secondly, I would like to extend my appreciation to my dissertation committee members: Dr. Alexandra S. Voloshina, Dr. Terence Sanger, and Dr. An H. Do for their invaluable insights and constructive feedback. I also would like to thank my qualifying exam committee: Dr. Kelli Sharp, Dr. Zoran Nenadic, and Dr. Michael McCarthy. Your expertise and guidance have been crucial in shaping my research and fostering my development as a competent scholar.

Thirdly, I am grateful to Dr. Toshihiro Kameda and Dr. Takashi Sakai for providing me with the opportunity to pursue my chosen path. Without your support, I would not have been able to undertake this degree.

I would also like to express my gratitude to all the participants in my research, who generously shared their time, experiences, and insights. Your contributions have been indispensable in deepening my understanding of the dissertation topic.

A special thanks goes to Justin B Rowe at Flint Rehabilitation Devices, and my collaborators: Diogo Schwerz de Lucena, Mina Ibrahim, Roman Yakunin, and Jathin Korrapati. Your collective efforts have significantly contributed to the iterative process of enhancing the efficiency of our HARCS project.

I am grateful to my colleagues in the BioRobotics Lab: Sumner Norman, Quentin Sanders, Merav Senesh, Veronica Swanson, Chris Johnson, Dylan Reinsdorf, Vicky Chan, Martí Comellas, Juan C Perez, Sangjoon Kim, Andria Farrens, Joshua Jones Macopson, Dongwon Kim, Caitlin Callaghan, Luis Garcia, and Guillem Cornella I Barba. I would like to particularly thank my first mentor in the BioRobotics Lab, Joan Lobo Prat. I'm deeply grateful for your guidance.

I would like to acknowledge the financial support provided by the National Institute on Disability, Independent Living, and Rehabilitation Research (NIDILRR grant number 90REGE0005-01-00). I am genuinely appreciative of this funding.

Finally, my heartfelt gratitude goes to my family - Takasuke, Kanae, Reiko, and Yoshiho - for their unwavering love and support throughout this demanding yet fulfilling journey. Without your support, completing this research would have been impossible.

To everyone who has supported and guided me along the way, thank you from the bottom of my heart.

# 謝辞

David に心より感謝申し上げます。私を研究室に受け入れてご指導いただき、ありがとうございました。

また、これまでご指導くださった先生方に心より感謝申し上げます。特に筑波大学の亀田敏弘先生と成蹊大学の酒井孝先生におかれましては、私の留学に際して多大なる援助をいただきました。また、本研究において多くの先生方より実り多きご指導とご助言をいただきましたこと、改めてここに深くお礼を申し上げます。

本研究を進めるにあたっては、共同研究者の皆様のご協力が不可欠でした。皆様のおかげで研究を進めることができました。また、実験に参加してくださった皆様に感謝申し上げます。

研究資金の援助をいただいた National Institute on Disability, Independent Living, and Rehabilitation Research (NIDILRR) に対し、深く感謝いたします。この支援がなければ、研究を進めることはできませんでした。

そして、この研究を取り組むことができたのは、両親と兄弟の支えがあったからこそです。私を支えてくれた家族に、心から感謝申し上げます。

本博士課程の研究を通してリハビリテーション、およびヘルスケア研究の意義深さを深く感じ、テクノロジーを通じて人々の生活を豊かにすることに貢献したいという想いがより強くなりました。これからの人生で、皆様に恩返しできるよう励みます。

沖田周祐

# VITA

**Shusuke Okita**

## EDUCATION

2023 Ph.D. in Mechanical and Aerospace Engineering, University of California, Irvine

2021 M.S. in Mechanical and Aerospace Engineering, University of California, Irvine

2016 B.A. in Mechanical Engineering, Seikei University

## PROFESSIONAL EXPERIENCES

2018 Part-time Intern, Cyberdyne Inc, Tsukuba, Ibaraki, Japan

2017-18 Visiting Scholar, University of California, Irvine, California, USA

## PUBLICATIONS

- **Shusuke Okita**, Diogo Schwerz de Lucena, Vicky Chan, and David J. Reinkensmeyer. 2023. (In Preparation) "Movement diversity and complexity increase as arm impairment decreases after stroke"

- **Shusuke Okita**, Roman Yakunin, Jathin Korrapati, Mina Ibrahim, Diogo Schwerz de Lucena, Vicky Chan, and David J. Reinkensmeyer. 2023. (Under Review) "Counting finger and wrist movements using only a wrist-worn, inertial measurement unit: Toward practical wearable sensing for hand-related healthcare applications " *Sensors*

- Schwerz de Lucena, Diogo, Justin B. Rowe, **Shusuke Okita**, Vicky Chan, Steven C. Cramer, and David J. Reinkensmeyer. 2022. "Providing Real-Time Wearable Feedback to Increase Hand Use after Stroke: A Randomized, Controlled Trial" *Sensors* 22, no. 18: 6938. https://doi.org/10.3390/s22186938

- Quentin Sanders, **Shusuke Okita**, Joan Lobo-Prat, Diogo Schwerz de Lucena, Brendan. W. Smith and David J. Reinkensmeyer, "Design and Control of a Novel Grip Amplifier to Support Pinch Grip with a Minimal Soft Hand Exoskeleton," *2018 7th IEEE International Conference on Biomedical Robotics and Biomechatronics*

*(Biorob)*, Enschede, Netherlands, 2018, pp. 1089-1094, doi: 10.1109/BIOROB.2018.8487918.

# ABSTRACT OF THE DISSERTATION

Novel Algorithms for Improving Wearable Feedback after Stroke

by

Shusuke Okita

Doctor of Philosophy in Mechanical and Aerospace Engineering

University of California, Irvine, 2023

Professor David J. Reinkensmeyer, Chair

Stroke is a leading cause of disability, with over 80% of patients experiencing chronic upper extremity (UE) impairments that impede daily activities and degrade quality of life. While wearable sensors in the form of step counters have found application in stroke rehabilitation, effective home rehabilitation of the UE using wearable sensors remains an open question. This dissertation focuses on three research questions in UE rehabilitation using a wrist-worn wearable inertial measurement unit (IMU): (1) identifying hand movements, (2) assessing the quality of movement experience (QOME) in daily life, and (3) identifying ways to improve UE QOME with wearable feedback.

We begin by reviewing existing wearable sensor technologies for at-home rehabilitation, setting the stage for our exploration of hand movement identification and arm movement quality assessment. We then introduce a spectrogram-based convolutional neural network (CNN) algorithm for hand movement recognition using a single wrist-worn inertial measurement unit (IMU). Our working hypothesis was that we

could use machine learning to identify active flexion and extension of the fingers/wrist based on the vibrational patterns produced at the distal end of the forearm. Using wrist-worn IMU recordings from 22 individuals with a stroke, we found we could identify the occurrence of finger/wrist movements with approximately 75% accuracy. Thus, ringless sensing of finger/wrist movement occurrence is feasible using wrist-worn IMUs, opening up new avenues for hand-related healthcare applications.

Subsequently, we posited that it may be more effective to encourage an increase in beneficial patterns of movement (i.e. QOME), rather than simply the overall amount of movement. As a first step toward this goal, we sought to identify statistical characteristics of daily arm movements that become more prominent as arm impairment decreases. Using the same data set as for the hand movement study, we identified several measures that increased as UE Fugl-Meyer (UEFM) score increased: forearm speed, forearm postural diversity (quantified by kurtosis of the tilt-angle), and forearm postural complexity (quantified by sample entropy of tilt angle). Dividing participants into severe, moderate, and mild impairment groups, we found that forearm postural diversity and complexity were best able to distinguish the groups (Cohen's D = 1.1, and 0.99, respectively) and were also the best subset of predictors for UEFM score. Based on these findings and a large body of research in motor learning that indicates the importance of challenging and variable movement practice, we posit that encouraging people to achieve more forearm postural diversity and complexity will improve QOME and therefore will be therapeutically beneficial.

Finally, we sought to identify practical ways to improve QOME with wearable feedback. In an experiment with unimpaired individuals, we examined specific exercises, chosen from a candidate list of common exercise activities, to find which exercises create a high level of movement diversity and complexity. Engaging in conventional rehabilitation therapy exercises created high values for forearm postural diversity but not complexity, as measured with a wrist-worn IMU. Playing the card game Speed and exercising with a commercial, gamified, home exercise sensor system produced the highest values for both postural diversity and complexity. Then, we designed an implementation for providing real-time, wearable feedback based on diversity and complexity, working toward a randomized controlled trial to assess the efficacy of QOME feedback compared to quantity feedback alone. We introduced the concept of "Quality Time" to measure the amount of time users perform complex, diverse movements. We propose an adaptive goal-setting strategy based on clinical scores of UE impairment and the statistics of complexity and diversity measured across a wide range of persons with stroke.

# Chapter 1. Introduction

This chapter reviews terminology and concepts relevant to the development of novel wearable sensing technology for upper extremity (UE) rehabilitation after stroke. Firstly, we review what constitutes effective rehabilitation for the UE after stroke, with a particular focus on the importance of movement dosage. Secondly, we review strategies for assessing UE function and measuring UE activities of people after stroke, with a focus on the emergence of wearable sensors. Finally, we discuss how wearable sensors might be used to improve neurorehabilitation. The chapter concludes by previewing the major contributions of the Dissertation.

## *1.1. What constitutes effective neurorehabilitation for the UE after stroke?*

Stroke is the fourth leading cause of death across the world [1]. It is expected that one out of four people may experience a stroke in their lifetime globally [2], and about 80% of acute stroke patients experience motor impairments [3]. Stroke occurs due to a blockage or a rupture of blood vessels [4], typically leading to impairment of descending neural pathways (i.e., the efferent neuronal pathways) [5] as well as ascending pathways (i.e. the afferent neuronal pathways) [6]. The damage to the central nervous system (CNS) impairs sensorimotor control for both the lower extremity (LE) and the UE. In this dissertation, we focus on remediation of UE impairments, which are characterized by limited dexterity due to 1) insufficient activation of agonist muscles; and 2) inappropriate activation of antagonistic muscles [7]. This muscle weakness and incoordination consequently reduces quality of life [8] and contributes to post-stroke depression [9]. Indeed, over 80% of persons who have experience a stroke express a

desire to improve UE performance in daily life [10], suggesting a strong need for more effective UE rehabilitation.

Although effective neurorehabilitation is crucial for restoring UE function, it still remains unclear which interventions or design specifications contribute to successful UE rehabilitation, particularly in the home environment [11]. Maier et al. identified 15 principles for effective neurorehabilitation [12]: massed practice, spaced practice, dosage, task-specific practice, goal-oriented practice, variable practice, increasing difficulty, multisensory stimulation, rhythmic cueing, explicit feedback/knowledge of results, implicit feedback/knowledge of performance, modulate effector selection, action observation/embodied practice, motor imagery, and social interaction.

Of the principles suggested by Maier et al., the number of movement repetitions (i.e., dosage) is often considered one of the most important components for effective UE rehabilitation, although attempts to validate this principle have been inconsistent [13], [14]. In one of the clearest findings of a dose-response effect of therapy, Winstein et al. recently observed a linear relationship between improvements of the Motor Activity Log (MAL) and the amount of therapy [15]. The MAL is a self-report scale for which individuals rate how much and how well they are using their impaired limb from 0 to 5. They split 41 chronic participants into four groups with varying duration of arm therapy from 0, 15, 30 and 60 hours. They found a gain of 0.92 in MAL for quality of movement in the 60 hours group compared to the 0 hours group.

The dose-response effect of training is also supported by work in animal models of stroke. Jeffers et al. used the Montoya stair-case test [16] to compare the relationship between the number of reach/grasp movements practiced and the skilled

motor recovery gain in this task for rats with an induced stroke [17]. They found that the number of practice movement repetitions was nonlinearly related to skilled motor recovery ($P<0.001$, $N=49$). Of particular interest here is the “hockey-stick” nature of the dose-response relationship they found. Small amounts of motor practice produced small amounts of additional recovery up until a threshold of about 300 practice movements per day. After that, additional amounts of practice were associated with linearly increasing skilled motor recovery.

The relationship between dosage to the restoration of UE function has been suggested to be explained by the Threshold Hypothesis, which states that UE therapy becomes less effective if you don’t practice enough number of movements above a threshold. Critical to the Threshold Hypothesis is the idea of feedback loops related to dosage that have been called the “virtuous cycle” and “vicious cycle” [18]. The former is identified as the cycle where arm-use promotes the increase of UE function, and better UE function in turn promotes arm-use. The latter refers to the cycle where non-use of the UE leads to poor function of the UE, and the poor UE function promotes arm disuse. The virtuous cycle and the vicious cycle were first proposed in the simulation [19], [20], and were verified with clinical experiments [15]. Due to the vicious cycle, subjects experience “rehabilitation in vain”, the state that the arm therapy is not effective in promoting recovery [19], [20]. Thus, it is encouraged to provide a sufficient dosage over a threshold for effective rehabilitation.

Previous research has shown that patients with hemiparesis after stroke may not receive enough movement practice in physical therapy (PT) and occupational therapy (OT) sessions [21]. Lang et al. found that treatment sessions averaged 36 minutes per

session, with an average of 39 repetitions of active-exercise movements, 34 repetitions of passive-exercise movements, and 12 repetitions of purposeful UE movements. This falls below the expectations required by the Threshold Hypothesis, suggesting that additional practice outside of therapy sessions is necessary to prevent ineffective rehabilitation. Therefore, it is encouraged to explore new ways to increase the number of repetitions for better recovery outside the clinic.

Another important consideration is that multiple studies suggest there is a discrepancy between the clinical assessment in the lab and the performance of daily life activities, as measured using wrist-worn sensors [22]–[24]. For instance, a recent study by Lang et al. followed 138 people receiving outpatient services at 5 rehab clinics in the US [25]. They found that 58% of people were able to improve their ability to complete upper limb and walking activities in the clinic, but did not see improvements in their movement performance outside of the clinic when using wearable sensors. They monitored the increase of the Action Research Arm Test (ARAT) score in the clinic, and the increase of the arm use ratio, defined as the ratio of impaired arm use to unimpaired arm use in daily life. They computed the arm use based on the duration of arm movements quantified by bilateral wrist sensors. They reported that stroke subjects didn't increase the arm use although they increased significantly in ARAT score from the baseline to the post therapy, meaning the improvement of the clinical score did not directly carry over to arm use in daily life [25].

In summary, stroke is a leading cause of motor impairment worldwide, particularly in UE function. Effective neurorehabilitation is essential for restoring UE function, with dosage, or the number of movement repetitions, being a crucial factor for

successful rehabilitation. The threshold hypothesis emphasizes the need for sufficient practice above a certain threshold to avoid rehabilitation in vain [19]. However, existing physical and occupational therapy sessions may not provide enough movement practice [21], indicating the necessity to explore alternative methods to increase the number of repetitions outside the clinic. Furthermore, a discrepancy between clinical assessment and daily life activities has been observed, suggesting that improvements in clinical scores may not directly translate to increased arm use in daily life [25]. This highlights the importance of developing novel approaches to enhance the transferability of UE rehabilitation from clinical settings to everyday life and underscores the need for more extensive research into effective strategies for optimizing post-stroke recovery.

*1.2. Assessing UE function and measuring UE activities of people after stroke*

Considering the pivotal role dosage plays in the effective recovery of UE function, it is crucial to accurately assess patients' motor impairments and measure UL activities. Several popular clinical tests are used to evaluate UE function, including the UE Fugl-Meyer (UEFM) test [26], Wolf-Motor Function Test (WMFT) [27], [28], and Action Research Arm Test (ARAT) [29], [30]. The UEFM comprises 33 test movements related to arm and hand dexterity, which are scored 0, 1, or 2 and then summed to obtain a score out of 66. The WMFT consists of 21 tasks involving functional movements [27], while the ARAT consists of 19 tasks assessing arm function using a kit [30].

To detect and measure UE movements inside and outside clinical settings, two approaches can be employed: (1) the optical sensor approach, such as optical motion capture systems [31], and (2) the inertial sensor-based approach, such as inertial

measurement units (IMUs). The optical approach identifies the kinematic features of human movements by employing optical sensors primarily placed in the laboratory settings. While the optical sensor based approach is predominantly used in clinical or laboratory environments, there have been studies using cameras to monitor activities of daily living (ADL); for instance, Likitlersuang et al. developed a computer vision that identifies subjects' impairment level based on a camera placed on a subject's chest, hanging from the neck [32]. They discovered a moderate but significant correlation between manual-labeling and the proposed approach, suggesting this method could serve as an alternative to manual labeling. On the other hand, inertial sensor-based approaches use inertial sensors that can be attached to different parts of the body to measure acceleration, angular velocity, and magnetic field strength. They are not affected by external factors such as lighting conditions or reflective surfaces [33], although they are affected by sensor drift. The degree of accuracy and reliability depends on the site of attachment.

In this dissertation, we focus on the use of inertial sensors for monitoring UE activities. Inertial sensors are versatile and can be utilized in various environments, making them an ideal choice for detecting and measuring UE movements. They are the primary sensor used for smart watches and activity trackers. They offer several advantages over optical sensors, such as being unaffected by lighting conditions or reflective surfaces and potentially providing accurate and reliable measurements depending on the attachment site.

The potential of wrist-worn sensors for monitoring UE activities has garnered significant attention in recent years. Research studies have provided valuable, new

insights into UE rehabilitation, since as the discrepancy between capacity and performance reviewed above, and both researchers and therapists have noted the potential to promote patient engagement in the rehabilitation process [34] . By leveraging the capabilities of these unobtrusive devices, therapists and patients alike can potentially access valuable information regarding movement patterns, progress, and areas for improvement.

Wearable sensing is a rapidly growing field in the healthcare market [35], with estimates projecting an increase from $36.34 billion in 2020 to $114.36 billion in 2028. As the market expands, wrist-worn sensors are suggested to modulate biomarkers related to health outcomes. For example, monitoring the number of steps taken through wearable sensors can help a person lower their body mass index (BMI) [36], and blood pressure [36] by providing insights into their walking habits.

Several commercially available wrist-worn devices have been developed or applied to facilitate UE rehabilitation at home. The MiGo (Flint Rehabilitation Devices, USA) [37] is a device available for research purposes that uses a six-degrees-of-freedom (DOF) IMU to track upper extremity movement. It has an accelerometer range of ± 2 G and a gyroscope range of ± 500 degrees per second with a 16-bit resolution. It is controlled by a system-on-a-chip (NRF52) with an ARM Cortex M4 CPU and a 2.4gHz radio. It can stream data at 100Hz or track upper extremity active time in 15-minute intervals for up to 37 hours. It can also be connected to a cellular gateway for extended logging range and to send data to a HIPAA compliant server. It has an OLED display and a push-button and is powered by a 90mAh LiPo battery.

ARYS™ me (Tyromotion GmbH, Graz, Austria) [38] is a discreet arm tracker and smartphone app system that helps individuals with arm impairments track and improve their arm movements through everyday activities. The wristband sensors measure all arm activities performed and the app provides a fun and motivating way to track progress, set daily goals and challenges, and receive reminders to move regularly. Additionally, the app includes a pedometer and the Tree of Recovery™ feature, which visualizes the progress of therapy over time. The system also includes a professional version, ARYS™ pro, specifically for use in clinical settings, which allows therapists to track and adjust therapy plans based on the patient's frequency of movement and progress. Both versions of ARYS™ support communication with patients, relatives, and insurance providers and can be used for both inpatient and home therapy. The software allows for objective therapy planning and coordination, making progress and limitations visible at a glance.

Armeo Senso (Hocoma, Switzerland) is a sensor-based solution for arm function recovery that is mobile, portable, and is designed to have an intuitive user interface and workflow. It comprises three sensors - two for the arm and one for the chest - a hand module, and a software program with game-like exercises that provide highly intensive yet motivating therapy. A multicentric, assessor-blinded, randomized controlled trial used the Armeo Senso virtual reality rehabilitation system to train 37 first-time stroke patients and found that reward during arm training reduced impairment and increased activity [39]. The device allows patients to train on their own or with very little supervision, which increases the number of repetitions they can perform.

Three other devices available for monitoring UE activity, primarily for research purposes are the Manumeter [40], [41], Actigraph GT3X (ActiGraph LLC, USA) [42], and Zurich-move. The Manumeter is a wrist-worn device developed in the UCI Biorobotics laboratory, featuring four magnetometers on its corners, allowing users to recognize hand movements based on magnetic field changes produced by a magnetic ring on the finger [40], [41]. The Actigraph GT3X contains a three-axis accelerometer to record arm movement data that has been used in a wide range of research studies [42]. The Zurich-move is another research-oriented device that combines accelerometry and gyroscope measurements to obtain detailed information on arm movement and function. By utilizing these research-focused devices, investigators have sought to gain insight into the effectiveness of various rehabilitation strategies and refine their approaches to maximize patient outcomes in the context of upper limb recovery.

*1.3. How could wearable sensors be used to promote effective neurorehabilitation?*

The effectiveness of upper extremity (UE) rehabilitation using wearable sensors, particularly wrist-worn devices, remains an open question. Previous research by Brenann et al. has suggested that there are four key feedback components for digital biofeedback systems, such as wrist-worn devices [43]:

(1) Content: Content refers to kinematic features provided as feedback to subjects. This can be categorized into knowledge of results (KR), and knowledge of performance (KP). KR refers to quantitative outcomes obtained from stroke subjects, and measured by wearable sensors, such as the number of movement

repetition. KP refers to information about the execution of a movement, like movement trajectory and force patterns over time [44].

(2) Timing (Concurrent or Terminal): Feedback timing can be either concurrent (delivered during exercise execution) or terminal (delivered after exercise execution). Wearable sensors can provide real-time feedback, delayed feedback, or a combination of both.

(3) Feedback Mode (Audio, Visual, and Haptic): Wrist-worn devices for UE rehabilitation can provide auditory, visual, and vibrotactile feedback [45]. Auditory feedback typically informs subjects if they have achieved a targeted movement. Visual feedback, often displayed on the device, can show number of movements completed as a progress toward daily goal.

(4) Frequency (Constant, Reduced, Fading): Feedback frequency can be categorized as “constant”, “reduced”, “fading”. Providing reminders in an appropriate manner can powerfully modulate motor learning outcomes [46].

In terms of content, as reviewed above, it is desirable for people after stroke to receive a large dosage of UE movement practice during rehabilitation therapy. To track how much movement practice has been done to reach a goal in daily life with wrist-worn device, it is helpful to develop algorithms that count UE movement. The primary approach that has been tried to date is to detect when the integrated magnitude of wrist acceleration exceeds a threshold, and to designate that crossing a “activity count” [47]. Most research studies in wearable sensing for UE stroke rehabilitation have used such activity counts as their primary metric [34].

A criticism of the activity-count approach is that it potentially does not distinguish generalized UE movement, since as arm swing during walking, from functional UE use [40]. Functional UE use typically requires use of the hand, and thus it would be desirable to develop a wearable sensor for measuring finger movement as opposed to generalized UE activity counts. Our laboratory previously developed sensors and algorithms that identify the number of finger/wrist movements a person makes throughout the day by detecting rotation of a magnetic field generated from a magnetic ring worn on the fingers using an array of magnetometers at the wrist. We ultimately achieved 75-80% accuracy in a controlled environment [41], [48]. However, this approach requires that users wear a magnetic ring, which in some cases users found bothersome because of the propensity of the ring to be attracted to metal objects like car doors or cooking ware. Other approaches to measuring finger movement require that individuals wear both a ring with an embedded accelerometer and a wrist accelerometer [49]. It would be desirable to be able to detect finger movement using wrist-worn IMUs, which are the most widely available commercial wearable sensing devices. Thus, an unsolved technical challenge is determining whether estimation of finger movement is possible using only IMU information obtained from the wrist.

In terms of timing, the most effective timing intervals for feedback in digital biofeedback systems depends on the individual user and the specific exercise being performed. Feedback timing can be concurrent, which means that feedback is given in real-time, or terminal, which means that feedback is given after the exercise is completed. Ling et al. used concurrent feedback in a digital biofeedback system for lower limb exercises and balance training, where a human avatar simulating the user's

movement performed a program of games with gamification feedback elements [50]. Spina et al. used both audio and visual concurrent feedback in a digital biofeedback system for balance training [51]. In general, concurrent feedback is more effective for motor learning and engagement with rehabilitation exercises [43]. However, the frequency of feedback should be adjusted based on the user's skill level and the difficulty of the exercise [43].

In terms of feedback mode, vibrotactile feedback provides sensory information using the afferent neural pathways with mechanoreceptors, such as Pacinian corpuscles and Ruffini endings, to the CNS [52], and has the potential to promote brain plasticity. Seim et al. recently developed a vibrotactile stimulation (VTS) glove that provides vibrotactile stimulation to the users through the skin [53], [54]. 16 chronic stroke survivors wore the VTS glove on the impaired hand for at least three hours outside the clinic to see if vibration feedback in daily life promotes UE recovery. They showed a statistically significant increase in Semmes-Weinstein monofilament exam results. As another example of vibrotactile feedback, Eagleman et al. investigated the ability of deaf and hard of hearing participants to identify sounds using spatiotemporal patterns of vibration on the skin of the wrist [55]. The subjects wore a vibrating wristband that converts sound to vibration. The wristband was created by Neosensory (Palo Alto, California), allowing users to "feel" sound through dynamic patterns of vibrations on their skin. The vibrating wristband and app offered non-invasive means of improving outcomes for hearing-impaired listeners. With the wristband, participants were able to identify up to 95% and on average 70% of the stimuli in a three-alternative forced choice task. Performance improved significantly over a month. Younger

participants tended to score better. The study also tested pattern discrimination, where participants determined whether the word presented was the same or different from the previous word. Performance was best with non-minimal pairs and could reach 100%. The study demonstrates the specific potential for wearable sensory substitution technology for the deaf and hard of hearing communities, but more generally illustrates how information content can be delivered through vibration.

In terms of feedback frequency, Wei et al. checked if subjects increase the amount of arm use by reminder them of making a movement using wearable sensor [56]. This study split 12 chronic stroke participants and 15 healthy participants into experimental group, sham group, and control group. The sham group and experimental group wore a wrist worn device that had the ability to provide a reminder for making a movement, but the sham group didn't receive reminder feedback from the device. In contrast, every 10 minutes, the experimental group were required to make a movement. They found that the average movement amount was significantly higher in the experimental group, and that this group also improved arm function more, as measured with the Action Research Arm Test score. Further, they showed that there was a distinct hemodynamic response in the primary motor cortex (M1) and the dorsolateral prefrontal cortex (DLPFC) using fNIRS when subjects received a reminder to move in the laboratory environment [57].

Finally, an important factor that has not been investigated thoroughly for successful at-home stroke rehabilitation with wrist-worn devices is knowledge-of-performance based feedback, which is also called quality of movement feedback [58], [59]. High-quality movements in UE rehabilitation have been defined in different ways,

including movements that are smoother [60]–[62], faster [63], [64], that exhibit better arm-hand coordination [63], that do not rely on compensatory patterns such a leaning with the trunk [65], and that do not rely abnormal synergistic patterns of movement [66]. As I propose in this dissertation, an alternate definition of high-quality movements in UE rehabilitation could focus on the statistical properties of extended periods of movement throughout the day. It has been suggested that an exploration of a greater variety of movements could promote recovery [66]. Using a cumbersome wearable system, Wolpert et al. found that the range of motion of human arm movements is typically confined to a relatively small volume [67], implying even individuals without impairment only explore a limited set of movements in their daily activities. This may account for the observation that individuals do not normally improve their arm motor skill through normal daily activity: they need to perform activities that challenge their arm motor skill. This aligns with one of the most consistent and fundamental findings in the motor learning research: challenging, variable task practice is necessary to promote motor learning [46].

The theme of using exploration to promote UE rehabilitation was recently explored by Krakauer et al., who developed a robotic training system to encourage individuals with a stroke to perform novel movements using a neuro-animation environment [66]. Participants explored with making different gravity-supported UE movements to drive the playful motion of a virtual dolphin, making it, for example, jump out of the water and do a flip when they achieved a specific, desirable patterns of UE joint coordination.

However, to our knowledge, there have not yet been studies that have adopted this exploration approach using wearable sensing. As detailed below, we propose that wearable sensing could be used to encourage individuals with a stroke to make novel and exploratory movements in their daily life, potentially provoking motor skill development.

### *1.4. Overview of the Dissertation*

In this chapter, we reviewed what constitutes effective rehabilitation for the UE after stroke, highlighting the importance of movement dosage. We also reviewed strategies for assessing UE function and measuring UE activities of people after stroke, detailing several wearable sensors that are available for these purposes. Finally, we discussed how wearable sensors might be used to improve neurorehabilitation.

The primary goal of this dissertation is to develop novel algorithms that can be applied with wrist-worn IMUs to address two of the problems we identified in the review: (1) the need to non-obtrusively monitor the amount of hand movement that individuals perform throughout the day, (2) the need to encourage individuals to perform high-quality movements during daily life that promote UE skill learning.

In Chapter 2, we develop a machine learning approach for identifying finger movements based on information from a single, wrist-worn IMU. Traditionally, it is necessary to use two sources of information to identify isolated finger movement: a ring and a wrist sensor [48]. Here, we use a convolutional neural network (CNN) [68] to learn to identify the occurrence of finger movements in daily activities based on the

spectrogram of accelerations and angular velocities those movements produce at the wrist.

In Chapter 3, we posit that it may be more effective to encourage an increase in beneficial patterns of UE movement, rather than simply the overall amount of movement each day. We therefore sought to develop novel statistical measures of the quality of the movement experience (QOME) of the UE that individuals with a stroke experience in daily life. We propose several measures of QOME based on measuring the statistical properties of the movement speeds and forearm postures experienced throughout the day from a sample of 22 persons with a stroke who wore a wrist IMU at home. We show that movement diversity and complexity, computed by kurtosis and sample entropy, increase as motor impairment decreases following stroke. We therefore propose movement diversity and complexity as potential targets for wearable feedback to promote QOME.

In Chapter 4, we ask the question “What exercises are helpful in improving movement diversity and complexity?” We recruited 8 unimpaired individuals and evaluated a set of 12 therapeutic activities for postural diversity and complexity. Engaging in conventional rehabilitation therapy exercises created high values for forearm postural diversity but not complexity. Playing the card game Speed and exercising with a commercial home-exercise sensor-based game system produced the highest values for both postural diversity and complexity. These results suggest that individuals with a stroke could be encouraged to practice specific exercises to increase QOME.

In Chapter 5, we describe the implementation of real-time, embedded algorithms for QOME feedback using a wrist-worn IMU. The system incorporates algorithms for measuring sample entropy and kurtosis in order to measure the amount of “Quality Time” that wearers achieve during daily activities. We propose an adaptive threshold for movement quality scoring, based on the wearer’s UEFM score and the minimal detectable change for entropy and kurtosis. We also developed a robotic simulator of the forearm as a test bench to validate these algorithms.

In Chapter 6, we summarize the contributions of this dissertation and suggest directions for future research.

# Chapter 2. Counting finger and wrist movements using only a wrist-worn, inertial measurement unit: Toward practical wearable sensing for hand-related healthcare applications

## *2.1. Summary of the Chapter*

Having the ability to count finger and wrist movements throughout the day with a nonobtrusive, wearable sensor would be useful for hand-related healthcare applications, including monitoring and rehabilitation after stroke, carpal tunnel syndrome, or hand surgery. Previous approaches to achieving this goal have required the user to wear a ring that is magnetic or a ring that contains an inertial measurement unit (IMU). Here, we demonstrate that it is possible to identify the occurrence of finger and wrist flexion/extension movements with approximately 75% accuracy based on vibrations detected by a wrist-worn IMU, a device similar to commercial fitness trackers. We developed an approach we call “Hand Activity Recognition through using a Convolutional neural network with Spectrograms” (HARCS), that trains a CNN to recognize finger/wrist movement occurrence based on the velocity/acceleration spectrograms those movements create. We validated this approach in two ways. First, we trained a CNN based on wrist-worn IMU recordings obtained from twenty stroke survivors during daily life, where the occurrence of finger/wrist movement was labeled using a previously-validated algorithm called HAND that used data from a magnetic ring and a wrist-worn magnetometer array. The daily number of finger/wrist movements identified by HARCS had a strong positive correlation to the daily number identified by the HAND algorithm

($R^2$=0.76, $p < 0.001$). Second, we trained a CNN on an IMU dataset recorded from nine unimpaired volunteers who performed finger, wrist, and arm movements, where we labeled the finger and wrist movements using simultaneous data recorded from an optical motion capture system. In this case, HARCS identified the occurrence of finger/wrist movement with 75% accuracy. Ringless sensing of finger/wrist movement occurrence is feasible using wrist-worn IMUs, opening up new avenues for hand-related healthcare applications.

*2.2. Introduction*

The hand plays a critical role in daily function. A wide variety of conditions, including trauma to the hand, developmental disorders such as Autism spectrum disorder [69], and neurologic injuries such as stroke [70] and spinal cord injury [71], diminish effective hand use. Overuse of the hand can also cause pain and injury, such as in carpal tunnel syndrome [72]. Treatment for these conditions usually relies at least in part on achieving targeted levels of daily hand activity, with the goal of promoting recovery by gradually increasing use or avoiding injury by limiting use. However, currently, there are few, non-obtrusive wearable sensors for quantifying daily hand use. This limits the ability of clinicians and patients to understand if target hand use amounts are being met and to adapt treatment plans.

Currently, there are several promising wearable approaches to finger movement sensing, including: 1) wearing a camera, typically around the neck, and inferring hand activity using computer vision [32]; 2) wearing a ring with an inertial measurement unit (IMU) and inferring hand activity from motion of the ring [73]; and 3) wearing a magnetic ring and inferring hand activity based on the changes in magnetic fields sensed at the wrist

[70], [41], [40], [48]. We recently used the last approach to confirm the hypothesis that real-world upper extremity (UE) hand use increases only for stroke survivors who achieve a threshold level of UE functional capability [48]. Here, we explored an even less intrusive approach that is suitable for implementation with hardware available with many commercial activity trackers that are worn like watches.

Our working hypothesis was that we could use machine learning to identify the vibrational patterns produced at the distal end of the forearm by active flexion and extension of the fingers and wrist. Previous studies have shown that vibrations induced by tapping the forearm can be read out using sensors in an arm band, highlighting the fact that there is informational content in vibrations that propagate through the forearm [74]. Here, we studied whether finger/wrist movement produced vibrations at the wrist that could be used to identify the occurrence of finger/wrist movement. To this end, we propose a novel approach for hand activity recognition using a convolutional neural network with spectrograms, named Hand Activity Recognition through Convolutional neural network with Spectrograms (HARCS).

### *2.3. Methods*

#### *2.3.1. Wrist-worn sensors*

We used data from two wrist-worn sensors in this study, the Manumeter and the MiGo. Both devices are non-commercial devices developed in a collaboration with the company Flint Rehabilitation Devices, LLC (Irvine, CA, USA).

The Manumeter is a watch-like device with inertial and magnetic sensing capabilities, the latter of which we have studied extensively in several previous studies

[40], [41], [48], [70]. For this study, we made use of the six degrees of freedom (DOF) Inertial Measurement Unit (LSM6DSL) it incorporates; the range of the accelerometer was set to ± 4G, and of the gyroscope was set to ± 500 degrees per second, both with 16-bit resolution. The Manumeter is equipped with an ARM Cortex M4 CPU (NRF52, Nordic Semiconductor, Norway) and a real time clock (PFC2123) calculates the time and the date. It has a 4G flash memory (MT29F4G01ADAGDWB-IT:G TR) that records data obtained from the IMU. The sampling rate was set as 52.6 Hz, high enough for capturing human movements [75]. The Manumeter has an OLED display that can show information, such as the time of day, or the number of hand movements performed by a wearer. In previous studies, for counting finger/wrist movements, the user also wore a magnetic ring on the index finger. A magnetic array in the wrist unit sensed changes in the magnetic field produced by the ring (see [48]).

The MiGo is also a watch-like device with IMU sensing, but it does not have magnetic sensing. It uses the same IMU (LSM6DSL), but the accelerometer range was set to ± 2 G and gyroscope range set to ±500 degrees per second, both with a 16-bit resolution). It contains the same real-time clock and microcontroller as the Manumeter. An integrated 2.4gHz radio is used to stream data from the IMU. IMU data is read and pre-processed at 205Hz, then streamed at 100Hz using the enhanced shock-burst wireless protocol. We down-sampled IMU data to 52.6Hz offline using Scipy's resample function [76]. The MiGo also has an OLED display and a push-button available and is powered with a 90mAh LiPo battery.

To summarize, for all experiments we used the same IMU with the same sampling rate, and this was mostly achieved with the Manumeter. We only used the MiGo for seven-subjects for the Mocap-Lab Dataset, as described in more detail below.

*2.3.2. Experiments and Data Sets*

For this project we needed to obtain IMU recordings taken during UE movements that sometimes included finger/wrist movements, and we also needed to know when a finger/wrist movement occurred, in order to train the network and then validate its capability. We relied on three datasets, where we name the dataset based on: 1) the way we identified whether a finger/wrist movement occurred (i.e. via the Manumeter or motion capture); and 2) the location of collection (at home or in the laboratory). All experiments were approved by the UCI Institutional Review Board and subjects provided informed consent.

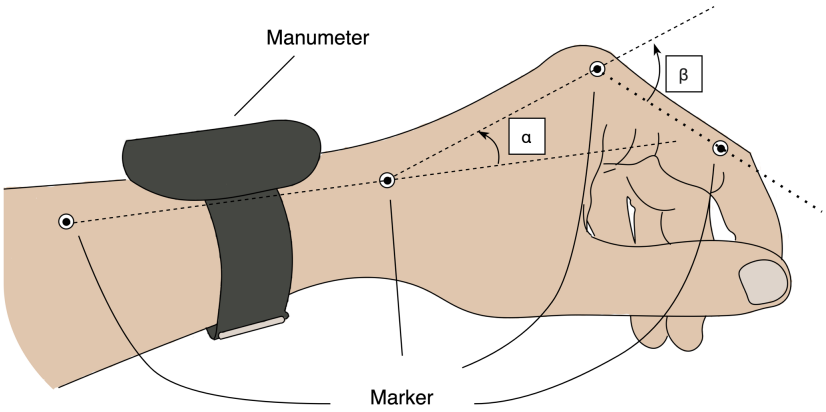
*Manumeter-Home Dataset:* For this dataset, we used IMU recordings from 20 persons with a stroke who wore the Manumeter at home, acquired as part of a previously-reported clinical trial on the feasibility and efficacy of hand count feedback [70]. We labeled movements using the HAND algorithm, which was previously developed in our laboratory [48]. The HAND algorithm recognizes hand movements based on fluctuations of the magnetic field produced from the magnetic ring on the wrist, detecting if its changes are over pre-set thresholds [77]. The algorithm identifies flexion-extension movements with an accuracy around 80%. Further details can be found in [48].

*Manumeter-Lab Dataset:* We also acquired IMU recordings in a laboratory-based experiment from the same subjects with a stroke who participated in generating the

Manumeter-Home Dataset. These subjects performed an exercise where they moved their hand or wrist a fixed number of times at a fixed pace by following a video prompt [70]. We also recruited an additional 7 unimpaired subjects to participate in a similar experiment, except these subjects performed an exercise where they moved their arm while keeping their hand still. Similar to the Manumeter-Home Dataset, we labeled finger/wrist movements using the HAND algorithm [48].

*Mocap-Lab Dataset:* To further validate the HARCS algorithm, we also acquired IMU recordings from nine, unimpaired, male volunteers who performed a series of structured UE movements while wearing either the Manumeter or the MiGo in the laboratory. For this experiment, we used an optical motion capture system (Phasespace, USA, nine cameras) to measure wrist and finger angles (Figure 1A). The participants performed six movements that involved various combinations of hand, wrist, and arm movement (Figure 1B). The subjects performed the same movement continuously over a 90 second period (at a self-selected rate, typically around 0.5-1 Hz), then rested for 15 seconds, before performing the next movement in the order shown in Figure 1B. We manually counted how many of each movement each subject performed. The dataset annotation process involved down-sampling Mocap data from 480 Hz to 52.6 Hz to match IMU data and calculating finger and wrist angles. Instantaneous angular rates were computed by differentiating these angles, and a window was labeled positive if a peak was found (see below for how windows were defined). Thresholds were set to match the number of peaks with the number of hand movements.

(A)




(B)

| Movement | Classification | Hand Movement Identification |
|---|---|---|
| Reaching Forward & Backward | Arm Movement | |
| Reaching Upward & Downward | Arm Movement | |
| Wrist Flexion & Extension | Hand Movement | Wrist Angle |
| Hand Open & Close | Hand Movement | Finger Angle |
| Reaching Forward & Backward + Hand Open & Close | Hand Movement + Arm Movement | Finger Angle |
| Reaching Forward & Backward + Wrist Flexion & Extension | Arm Movement + Hand Movement | Wrist Angle |

*Figure 1. (A) Marker placements for inferring the wrist and the finger angle. $\alpha$ represents the wrist angle. and $\beta$ represents the finger angle (i.e. metacarpophalangeal joint angle). Four markers were taped as shown in the figure to the wrist and finger to get $\alpha$ and $\beta$. (B) The list of movements in Mocap-Lab Dataset. Subjects performed 6 movements involving arm-only movement, hand-only movement, and hand & arm movement.*

### *2.3.3. Data Preprocessing*

We used 6 steps to annotate and preprocess datasets, and to train the network (Figure 2). We used 9 features in training the network: 3-axis acceleration, 3-axis angular velocity, and the 3-axis gravity direction. Note that we subtracted gravity from the acceleration, and obtained gravity direction through the Madgwick filter [78]. In Step 1, we annotated the datasets using either the motion capture system or the HAND algorithm. In Step 2, we split data into training and testing sets (see details in the "*Training the network*" section). Then, we generated windows with 150 samples (i.e., about 2.85 seconds of data),

which served as inputs for the training and validation datasets. For the training set, the windows were generated with a stride of 50 time-samples (about 0.95 seconds of data). We labeled a window as positive if at least one time-sample was designated as positive for hand movement within the last 50 time-samples of that window. We adopted this approach to provide the network with adequate context to determine if a hand movement has taken place; i.e. the network evaluated a window of 150 time-samples to predict if a hand movement took place in the last 50 time-samples in the window. As a result, for every 50 time-samples, we assigned a label of either positive or negative for hand movement occurrence, and defined those 50 time-samples, as well as the previous 100 time-samples as the "data-sample" for performing identification (see Figure 3 for an example of a data-sample). In Step 3, we converted the samples of raw sensor signals for each data-sample into spectrograms by applying a Short-Time Fourier Transform (STFT). In Step 4, we applied a non-linear transformation, the Box-Cox transform [79]–[82]) to the spectrogram variables for normalization, which we found improved network performance. During the generation of the training set, the parameter $\lambda$ for the Box-Cox transform was selected using the SciPy scalar optimizer [76]. After the normalization, we applied standardization by subtracting out the mean and dividing by the standard deviation. Further details are found in the Supplementary Material.

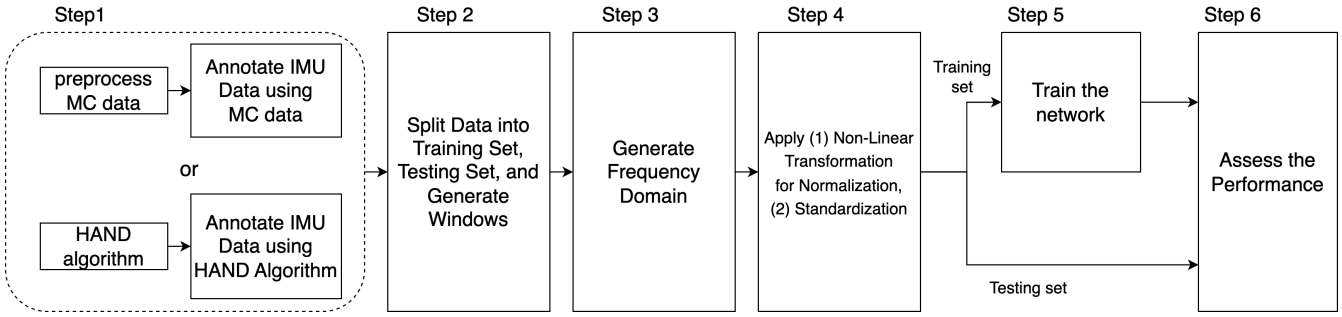


*Figure 2. 6 steps in training and assessing the neural network.*

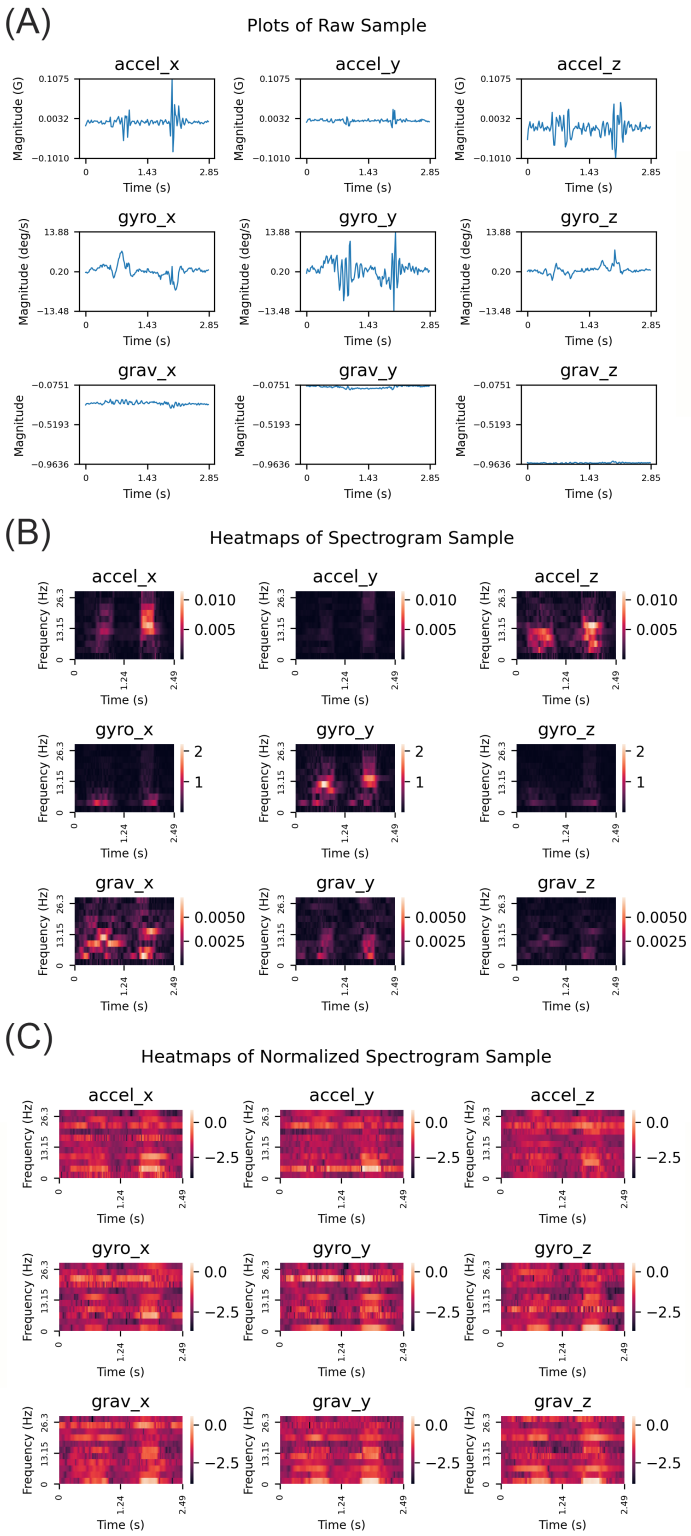



*Figure 3. An example of processing a data-sample where the HAND algorithm predicted there exists a hand movement. This sample contains 150 time-samples from the IMU. (A) Acceleration, gyro, and gravity vector measures (B) The heatmaps of the spectrograms*

*computed from this sample for each measurement. (C) The heatmaps of the normalized spectrograms.*

### *2.3.4. Convolutional Neural Network Design*

*In Step 5,* we trained a Convolutional Neural Network (CNN), a network architecture often used in image recognition tasks (Figure 4) [68]. Spectrogram classification can be effectively executed with CNNs [83], and there are various studies available on the application of spectrograms for classification tasks, specifically within the realms of voice [84] and human activity recognition [85]. The CNN we used consisted of 7-8 repeated convolutional layers with L2 regularization and dropout layers. The dropout layers were inserted to avoid overfitting [86]. A sigmoid activation function was used to output final predicted probabilities. For the loss function, we used a binary cross-entropy loss function with the Adam optimizer [87].

Each data-sample was formatted as 131 x 9 x 11, which is a transformed version of the original 150 time samples data with 9 sensor variables. The Short-Time Fourier Transform (STFT) was applied using a window size of 20, shifting one-time sample at a time. This analysis produced 131 windows from the initial 150 time-samples.

Tensorflow and Keras 2.0 were used to implement the CNN [88], [89]. We manually tuned the following parameters depending on a network: (1) the number of convolution layers; (2) the convolution kernel size; (3) the number of convolution filters; (4) whether layer norm layers are present; (5) l2 regularization lambda. We trained the network using either the Mocap-Lab Dataset or Manumeter-Home Dataset. A summary of the parameters in each case is provided in the Supplementary Material.

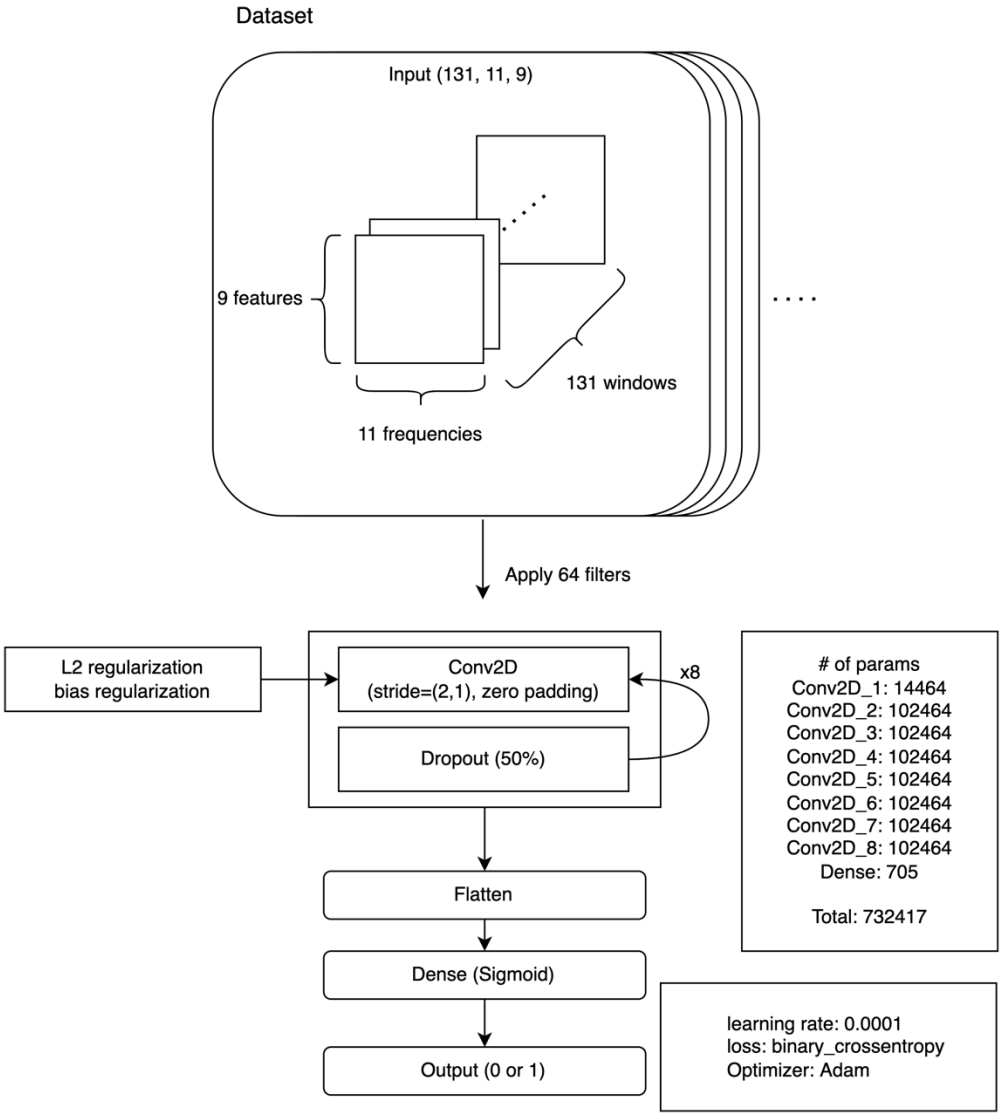



*Figure 4. An example of the CNN architecture trained by Manumeter-Home Dataset.*

### *2.3.5. Training the Network*

We trained the CNN with the different IMU datasets, which, to review, were taken in different settings and with finger/wrist movements identified in different ways. We asked the following questions:

*Q1.* How well can we identify in-the-wild hand movements made by people with varying levels of hand impairment due to stroke? Our largest dataset was the one obtained across a day during home activities by persons with a stroke, i.e. the Manumeter-Home Dataset. We asked: can a CNN trained on IMU recordings obtained from daily use of the UE at home by persons with stroke recognize when finger/wrist movement occurred? To answer this question, we used leave-one category-out-cross-validation (LOOCV) based on subjects' impairment levels. We trained HARCS by splitting the data into training and testing sets according to the subjects' Fugl-Meyer Upper Extremity (UEFM) scores. The UEFM score is a clinical scale ranging from 0 (severe impairment) to a maximum of 66 (no impairment), assessing motor function and joint functioning in individuals with post-stroke [26]. For instance, we trained HARCS using subjects' data in $30 \leqslant$ UEFM score $< 66$, when we assessed subjects' data in the range of UEFM score $< 30$. We labeled the occurrence of a finger/wrist movement using the previously-validated HAND algorithm implemented with the magnetic sensing capability of the Manumeter, which is about 80% accurate [48]. In addition to the LOOCV method based on subjects' impairment level, we also implemented an alternative approach to evaluate the performance of the CNN, which we called random 5-fold cross validation grouped by participants. In this method, we randomly divided the dataset, allocating 80% of the subjects for training and 20% for testing, without taking into account the subjects' UEFM scores. This approach aimed to assess the robustness and generalizability of our CNN model when mixing a diverse range of hand impairments due to stroke into training data. We performed 6 iterations of this

random fold process, each time creating a new training and testing data split. We also compared HARCS to three other machine learning approaches: the support vector machine (SVM), k-nearest neighbor (kNN), and a multi-layer perceptron, following the approaches by [90], [91], and implementing the algorithms in scikit-learn [92]. See the Supplemental Material for parameter selection for the other approaches. We also studied how the identification accuracy depended on finger-wrist movement speed. To do this, we calculated the mean of the acceleration amplitude across the 150 data points in each window. After we created a validation set, we split the data with respect to the amplitude of movements to compute a weighted confusion matrix in each range to average the effect of skewness. We averaged based on the number of actual positives and actual negatives.

*Q2.* Is HARCS Sensitive to Isolated Hand Movement?: A second question we asked was: Can a CNN trained using the Manumeter-Home Dataset accurately count hand movements when a person performs structured movements comprised of hand-only movement, and, conversely, not count movements comprised of arm-only movement? To answer these questions, we used the Manumeter-Lab Dataset. Since hand and arm movements often occur together in the "wild", our goal here was to understand the unique sensitivity of the algorithm to isolated hand movement, and, conversely, it's susceptibility to arm-only movement.

*Q3.* Identifying Structured Hand Movements with Accurate Labeling: Using the Manumeter to label movements introduces error because the HAND algorithm is about 80% accurate [48]. Therefore we asked: Can a CNN trained on movements counted accurately using motion capture recognize when a finger/wrist movement

occurred? We also varied the type of movement (hand only, arm only, and hand-arm together) to study accuracy for these movement types. To answer these questions, we trained and evaluated the CNN using the Mocap-Lab Dataset with LOOCV. We first removed one subject's data as the validation data set, and then trained the CNN on the data from the remaining eight subjects. We repeated the same process 9 times, using each subject's data for validation. We used the testing set to evaluate how well the CNN detected the presence or absence of finger/wrist movement for each of the six movements shown in Figure 1B. Moreover, for the Mocap-Lab Dataset, we implemented two distinct training approaches to evaluate the potential impact of different labeling methods on the network's performance. In the first approach, we only labeled hand-only movements as positives; that is, we labeled combined hand/arm movements as negative. This tested how well the network could identify when hand-only movements occurred, in isolation from arm movement. In the second approach, we labeled combined hand/arm movement as positives. This tested how well the network could identify when hand movement occurred, with or without arm movement.

*2.3.6. Statistical Analysis and Performance Analysis*

To characterize the network performance, we used Accuracy, Precision, Recall, and F1-Score, as defined in Equations 1-4, all widely used metrics [44], Additionally, we generated receiver operating characteristic (ROC) curves and associated Area Under the Curve (AUC) values to assess the performance of the model's binary predictions [94]. We

also computed the Pearson correlation to compare hand counts between HARCS and the HAND algorithm.

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive + False Positive}} \tag{1}$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive + False Negative}} \tag{2}$$

$$\text{F1 score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision + Recall}} \tag{3}$$

$$\text{Accuracy} = \frac{\text{True Positive + True Negative}}{\text{True Positive + False Positive + False Negative + True Negative}} \times 100 \tag{4}$$

*2.4. Results*

We tested the ability of HARCS to identify finger/wrist movement from wrist-worn IMU data sets obtained at home from people with hand impairment after a stroke and in the lab from both stroke and unimpaired subjects. HARCS uses a CNN to recognize finger/wrist movement occurrence based on the velocity/acceleration spectrograms those movements create.

*2.4.1. HARCS can identify unstructured hand movements of people with stroke in-the-wild across a wide range of impairment levels*

We divided the 20 stroke subjects into groups according to a standard clinical measure of UE movement ability, the UEFM Score. For reference, an UEFM score less than 20 means severe UE paresis, a score of 30-40 means hand function is just emerging; and a score of 66 means normal movement ability [95]. HARCS accuracy was 81% for the two subjects in the range of 0-20 in the UEFM score, and 74% for 15 subjects in the range 50-60; that is, the accuracy did not vary substantially with level of hand impairment (Figure 5A). Across all participants, using random 5-fold CV, the average accuracy was 77%. Further, the number of HARCS counts identified across a days' worth of wear time was strongly correlated with the number of HAND counts, which were the counts identified in the previous study that used additional information from a magnetic ring, instead of just IMU data ($r=0.874$, $p=0.00$, $R^2 = 0.763$) (Figure 5B).

(A)

| The Range of UEFM Score (Number of Subjects) | Accuracy (%) | AUC (%) |
|---|---|---|
| UEFM ≤ 20 (n=2) | 81.05 | 89.5 |
| 20 < UEFM ≤ 30 (n=5) | 78.69 | 87.17 |
| 30 < UEFM ≤ 40 (n=6) | 76.68 | 84.64 |
| 40 < UEFM ≤ 50 (n=3) | 80.61 | 87.32 |
| 50 < UEFM ≤ 60 (n=15) | 74.44 | 82.22 |
| Average (Random 5-fold CV) | 77.19 | 85.03 |

(B)

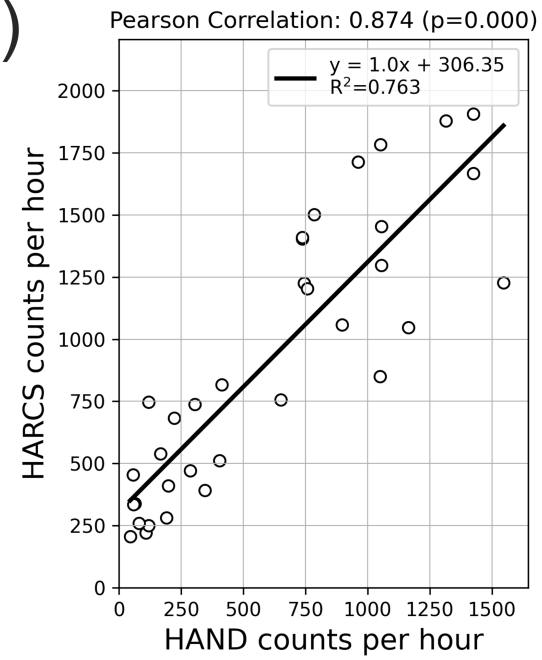



*Figure 5. (A) Accuracies and AUCs for HARCS-based identification of finger/wrist movement in-the-wild for 20 stroke subjects with varying UEFM scores. (B) The correlation between HARCS and HAND counts of finger/wrist movement occurrence, where HAND counts were produced in previous study using information from a magnetic ring.*

Using the same Manumeter-Home Dataset, we compared four machine learning approaches (see Supplemental Material for details). The various measures of network performance we examined were highest for HARCS, including the $R^2$ value for the regression between HARCS and HAND counts (Figure 5B).

*Table 1. Summary of performances by four different machine learning approaches.*

| | KNN | SVM | Perceptron | HARCS |
|---|---|---|---|---|
| Accuracy | 61.82 | 66.87 | 72.35 | 77.19 |
| F1 score | 44.22 | 61.37 | 70.71 | 77.98 |

| | | | | |
|---|---|---|---|---|
| Precision | 82.03 | 73.61 | 75.18 | 76.02 |
| Recall | 30.51 | 52.81 | 66.98 | 80.62 |
| $R^2$ | 0.648 | 0.31 | 0.533 | 0.763 |

Again using the same Manumeter-Home Dataset, we analyzed how the HARCS classification performance depended on the average speed of limb movement in the data-sample window (Figure 6). Accuracy and precision decreased with increasing average speed, while Recall increased. The decrease in precision could be attributed to an increase in false positives as speed increased (Figure 6F), while the increase in recall was due to a reduction in false negatives as speed increased (Figure 6G). Despite the balanced metrics, this outcome indicates that the network had a tendency to predict positive outcomes when the mean of accelerations was high. In other words, as the acceleration increases, the network is more likely to identify a movement as positive, leading to a higher rate of false positives and consequently lowering Precision. Conversely, the decrease in false negatives contributed to the increase in Recall, demonstrating that the network more effectively recognized actual positive cases.

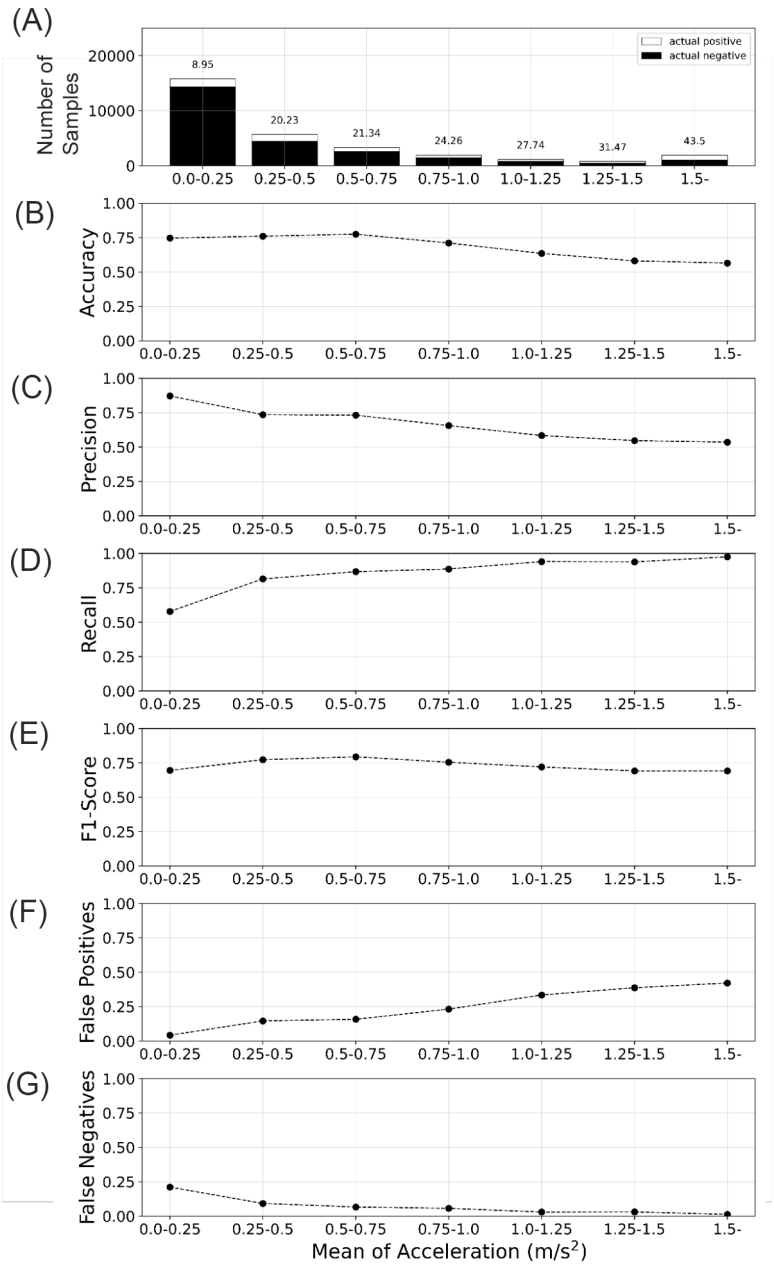



*Figure 6. The relationship between the network performance and the mean linear acceleration across the data-sample window of duration ~2.8 sec. In (A), the proportions of actual positives and total data-samples are shown on the top of each bar.*

*2.4.2. HARCS sensitivity to hand-only and arm-only movement*

After we trained HARCS on the Manumeter-Home Dataset, we tested its abilities on the data set from impaired and unimpaired subjects in which subjects performed hand-only or arm-only movements. This was of interest because hand and arm movements often occur naturally together in-the-wild, and we wanted to understand how sensitive HARCS is to isolated hand movement. HARCS was able to count hand-only movements, but underestimated the actual number of hand movements performed by both impaired and unimpaired subjects by 45% and 66% respectively. This performance was less accurate than the HAND algorithm, which overestimated the counts by 26 and 5%, respectively (Figure 7A and B). For the arm-only exercise, HARCS should not have counted any hand movements, but counted 36% of arm movements as having also had a hand movement, a false-positive rate that was again larger than HAND (Figure 7C).

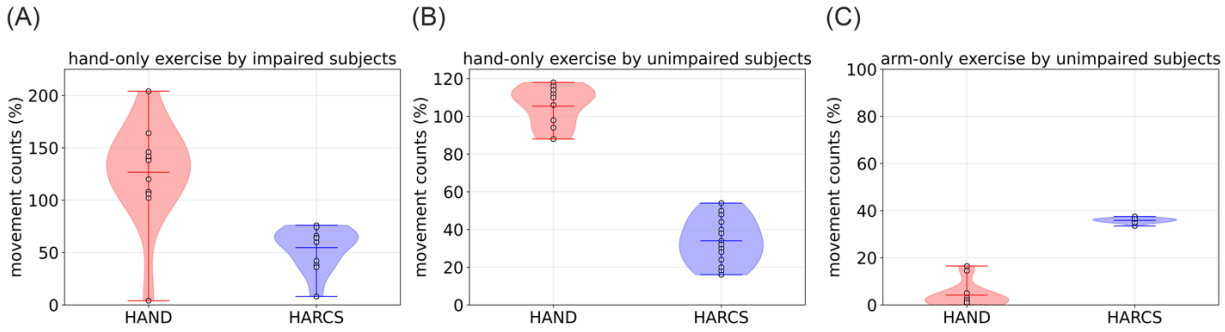



*Figure 7. Comparison of HARC algorithm counts versus HAND counts and the true number of hand movements for (A) hand-only exercise by persons with stroke (B) hand-only exercise by unimpaired subjects, (C) arm-only exercise by unimpaired subjects. For A and B, perfect counting would result in 100% movement counts. For C, perfect counting would result in 0% movement counts, since C is arm-only exercise. 100% represents 50 hand movements for hand-only exercise and 200 movements for arm-only exercise.*

*2.4.3. HARCS can identify structured hand movements, but accuracy depends on movement type*

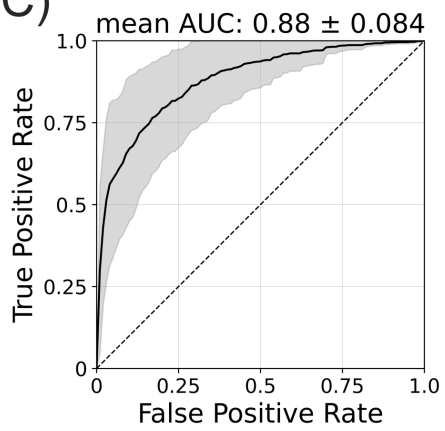
As a further test of the core idea of identifying hand movement from wrist IMU recordings, we performed a laboratory experiment that generated the Mocap-Lab Dataset. For this experiment, subjects repeatedly made different types of movements involving the hand and/or arm (Figure 1B). Hand movement occurrence was identified based on an optical motion capture system, removing any labeling error that resulted from depending on the HAND algorithm for labeling. The average accuracy for HARCS for the Mocap-Lab Dataset when combined hand/arm movements were treated as actual positives was 75%, but the accuracy varied depending on the type of movement that the subject performed (Figure 8A). HARCS performed well in identifying isolated finger/wrist movement (80 and 76% accuracy, see Figure 8A), and in ignoring vibrations created by arm movement only (81 and 91% accuracy). HARCS struggled to accurately identify finger/wrist movements combined with arm/hand movement (59 and 65% accuracy). Across all movement types, the average AUC was 0.88 (Figure 8C). In contrast, the average accuracy for HARCS for the Mocap-Lab Dataset when combined hand/arm movements were treated as negative was 92% on average, showing the ability of the network to more accurately identify hand-only movement (Figure 8B and D).

HARCS when combined
hand/arm movements were treated as actual positives

(A)

| | Average Accuracy ± One Standard Deviation |
|---|---|
| Reaching Forward & Backward | 81.63 ± 22.84 |
| Reaching Upward & Downward | 91.34 ± 22.11 |
| Wrist Flexion & Extension | 75.98 ± 26.78 |
| Hand Open & Close | 79.80 ± 31.88 |
| Reaching Forward & Backward + Hand Open & Close | 58.80 ± 29.36 |
| Reaching Forward & Backward + Wrist Flexion & Extension | 65.20 ± 29.76 |
| Average | 75.46 ± 28.20 |

(C)




HARCS when combined
hand/arm movements were treated as actual negatives

(B)

| | Average Accuracy ± One Standard Deviation |
|---|---|
| Reaching Forward & Backward | 98.86 ± 2.68 |
| Reaching Upward & Downward | 98.58 ± 2.32 |
| Wrist Flexion & Extension | 76.89 ± 26.90 |
| Hand Open & Close | 79.06 ± 31.48 |
| Reaching Forward & Backward + Hand Open & Close | 99.76 ± 0.72 |
| Reaching Forward & Backward + Wrist Flexion & Extension | 99.02 ± 2.94 |
| Average | 92.03 ± 19.06 |

(D)

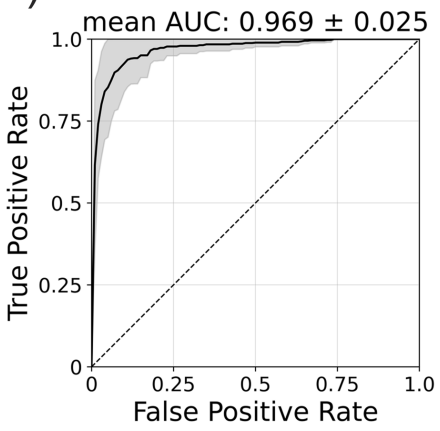



*Figure 8. (A, B) Average accuracy of HARCS when trained and tested on the Mocap-Lab Dataset obtained from nine unimpaired subjects. (C, D) The mean of ROC Curve (solid line) in which the shaded area represents ± 1 standard deviation. In A and C, combined hand/arm movements were treated as actual positive. In C and D, combined hand/arm movements were treated as actual negative.*

### *2.5. Discussion*

Our goal in this study was to determine if we could identify the occurrence of hand movements using only a wrist-worn IMU. The algorithm that we designed, called HARCS,

focuses on recognizing vibrational patterns produced by flexion and extension of the fingers and wrist. Using a dataset obtained from twenty stroke survivors during daily life, the daily number of finger/wrist movements identified by HARCS had a strong positive correlation with the number identified by a previously developed hand-movement recognition algorithm that used a magnetic ring and magnetometer. This was true across a wide range of hand impairment levels after stroke. Further experiments in the lab found HARCS has an accuracy of ~75% and was sensitive to the occurrence of hand-only movement, although it tended to undercount hand-only movement. These results demonstrate the feasibility of ringless sensing of hand movement using wrist-worn IMUs. We discuss the performance of HARCS as well as limitations and directions for future research.

*2.5.1. Performance Analysis*

In contrast to existing approaches of hand activity detection [48], [49], one of the advantages of HARCS is it doesn't require any other components or devices except a wrist-worn IMU. Thus, HARCS could potentially be deployed immediately with a wide range of existing smart watches. Encouragingly, hand counts produced by HARCS were strongly correlated with the HAND algorithm, which required additional hardware, including a magnetic ring and magnetometer array. This strong correlation was present even though the HAND algorithm is known to be susceptible to false positives due to stray magnetic fields in the environment [48], introducing noise in the labeling.

In the laboratory experiments, the overall recognition accuracy of ~75% was comparable to accuracy rates of some commercial step counters [96], [97], which have

been shown to be valuable for encouraging walking activity and improving body mass index and blood pressure, [36]. An analysis of 42 studies in real-world settings found that 58% of step trackers had more than 10% inaccuracy for step counts [97]. Step count error rates are even higher for impaired populations [98]. Despite not being perfectly accurate, HARCS may be reliable enough to encourage rehabilitation outcomes for stroke patients when used consistently. In fact, our previous study involving the Manumeter, which had an accuracy of about 80%, demonstrated that subjects increased their daily frequency of hand movements in response to the display feedback from the device [70].

Two key design features of HARCS are to convert IMU data to the frequency domain before using in classification, and to use a CNN network, an architecture that has been previously found useful for image recognition tasks [68]. Essentially, HARCS can be viewed as performing image recognition on the 9 spectrograms corresponding to the 9 types of features produced by finger, wrist, and arm movement.

In the first laboratory experiment, we found that HARCS was sensitive to isolated hand movement but undercounted it by ~30%. Further, HARCS assigned false positives at a rate of about 40% to arm-only movement. In this experiment, we trained the network using the in-the-wild data from the stroke subjects, which may account for some of the inaccuracy. As mentioned above, we labeled the in-the-wild data with the HAND algorithm, which is imperfect. Further, the movements types performed in the lab were only a small subset of the ones the network was trained on (i.e. daily life movements), and this may have biased the network in a suboptimal way for the laboratory tests. Finally, the testing data for arm-only movement were from unimpaired subjects, while the training data were from stroke subjects. Stroke subjects tend to move more slowly, and we found that the

network was sensitive to the average movement speed in the data-window. It may be possible to improve HARCS performance for specific applications by better matching training data to the application.

In the second laboratory experiment that used the motion capture system to accurately label data, we found that network performance was excellent when we trained the network to identify when hand-only movement occurred. Network performance decreased when we trained the network to identify when any type of hand movement occurred, including those occurring simultaneously with an arm movement. This highlights that there is a confounding nature to the vibrations produced by arm movement when trying to identify hand movement, as was also evident by the algorithm producing false positives for arm-only movement. This confound may present a fundamental limit to the accuracy possible with HARCS, although this remains an open question.

### *2.5.2. Limitations and Future Work*

A key limitation of the current work is that the in-the-wild data analysis relied on the magnetometer-based HAND algorithm for labeling, which we know from previous studies has some inaccuracies. It may be possible to improve performance by generating a more accurately labeled, in-the-wild dataset with a more obtrusive method, such as wearing instrumented gloves [41], wearable cameras [32], [99], or stretchable e-textile sensors [100], [101] to obtain a more accurate ground truth for annotating the dataset.

For future work, we aim to incorporate non-obtrusive hand movement sensing into home rehabilitation after stroke, allowing for real-time feedback and analysis of patient hand movements, an approach that already showed promise with a more cumbersome

magnetic sensing approach [70] . To achieve this goal, we will seek to embed HARCS into a wrist-worn sensor to provide real time hand activity recognition. Such a real-time version of HARCS could potentially be useful for other healthcare applications where monitoring the real-time use of the hand in daily life is relevant, including developmental disorders, spinal cord injury, hand trauma, and repetitive stress injuries.

# Chapter 3. Movement diversity and complexity increase as arm impairment decreases after stroke: Quality of movement experience as a possible target for wearable feedback

*3.1. Summary of the Chapter*

Over 80 percent of people who have experienced a stroke incur upper extremity (UE) impairment resulting in reduced arm use in daily life. A few studies have examined the use of wearable feedback of the quantity of arm movement in order to promote recovery, but with limited success. We posit that it may be more effective to encourage an increase in beneficial patterns of movement – i.e. the quality of the movement experience – rather than simply the overall amount of movement. As a first step toward this goal, here we sought to identify statistical characteristics of daily arm movements that become more prominent as arm impairment decreases, based on data obtained from a wrist IMU worn by 22 chronic stroke participants during their day. We identified several measures that increased as UE Fugl-Meyer (UEFM) score increased: forearm speed, forearm postural diversity (quantified by kurtosis of the tilt-angle), and forearm postural complexity (quantified by sample entropy of tilt angle). Dividing participants into severe, moderate, and mild impairment groups, we found that forearm postural diversity and complexity were best able to distinguish the groups (Cohen's D = 1.1, and 0.99, respectively) and were also the best subset of predictors for UEFM score. Based on these findings, we posit that encouraging people to achieve more forearm postural diversity and complexity might improve the quality of their movement experience and therefore might be therapeutically beneficial.

*3.2. Introduction*

Stroke is one of the most prevalent diseases across the world [1]. It is reported that 80 percent of people experience upper extremity (UE) motor impairments following stroke [102]. Due to these impairments, people rely more on their less-impaired UE for daily activities [103], although the activity level of both UEs decreases as the impairment of the paretic arm increases [104], [105]. This phenomenon relates to the concept of “learned non-use” [106] and may contribute to poor adherence to home exercise programs [107] and reduce gains from home rehabilitation [18].

Following on these concepts, Han et al. suggested that people following stroke can be categorized as either “users”, or “non-users” based on how much they use their impaired arm in daily activities [19]. “Users” are hypothesized to enter a “virtuous cycle”, in which they actively use their impaired arm, resulting in an improvement of arm function [18]. “Non-users”, in contrast, are hypothesized to enter a “vicious cycle”, in which their arm non-use leads to reduced arm function [18].

Escaping this vicious cycle is challenging due to multiple factors. Firstly, people generally don't begin to use their impaired arm voluntarily unless its function reaches at least half of its normal capacity [70], which aligns with the threshold hypothesis proposed by Schweighofer et al [19]. Secondly, even when individuals show improvement on clinical functional tests, such as the Action Research Arm Test (ARAT), this progress doesn't necessarily translate to increased arm use in daily life, indicating a "translation gap" between arm functionality and actual use [108]. For example, a recent study found that a majority of stroke patients improved the capacity for UE activity as they recovered;

however, they did not improve the actual performance of UE activity in daily life, as measured with a wearable sensor [109]. Further, patients typically spend limited time in therapeutic activities compared to non-therapeutic activities, with stroke patients spending only about 28% of their available time on therapeutic activities even in rehabilitation centers [110]. Finally, patients tend to overestimate the amount of their activity in home exercise programs [111].

We recently studied the potential of wearable feedback of amount of hand use to help address several of these issues. We developed the Manumeter, a wrist-worn magnetic array that senses the movement of a magnetic ring worn on a finger, indicating the number of hand movements on a display [40], [41], [48], [70]. We used the Manumeter to test the hypothesis that real-time, daily feedback could promote hand use in daily activities [48]. We found that providing three weeks of such feedback promoted a small but a statistically significant increase in the amount of hand movements. Further, UE movement ability, measured by the UE Fugl-Meyer (UEFM) Test and the Box and Block Test (BBT), modestly increased. Other studies of wearable feedback have also found modest benefits to providing wearable feedback of amount of hand use [112], [113], leaving open the question of whether there is a way to build on these results and increase the benefit of wearable feedback.

Here, we reason that people with a stroke not only use their UE less, but also have an impoverished daily movement experience because they don't use their UE for a variety of tasks. This idea of impoverishment can be considered to relate to the idea of "movement quality", but is somewhat different than how movement quality has typically been conceived in wearable sensor research. For instance, previous studies have found that

people after stroke exhibit decreased movement smoothness both during supervised movement in the clinic [60] and during daily activities [114]. Therefore, movement smoothness has been suggested as a measure of movement quality, or, alternately, measures such as increased movement speed and range of motion, or decreased curvature of reaching movements have been suggested [115]. However, these measures of movement quality focus on the quality of individual movements rather than the overall quality of the daily movement experience.

The idea of impoverishment we propose here relates to the idea that severely impaired stroke subjects appear “stuck” in stereotypical, abnormal synergistic movement patterns due to reduced capability of the corticospinal tract and/or the presence of spasticity [116]. If one only ever practices the same thing, it is difficult to improve in skill. Thus, we postulate that it is important for non-users to practice a more diverse set of challenging movements frequently throughout the day to enter a virtuous cycle that improves movement ability.

But what should this diverse set of more challenging movements look like, and how might it be sensed by a wrist-worn inertial measurement unit (IMU)? This study sought to identify statistical characteristics of daily arm movements, quantifiable with a wrist-worn IMU, that become more prominent as arm impairment decreases. We reasoned that better-recovered persons move in ways that would be advantageous for less well-recovered persons to attempt. Thus, we analyzed IMU data obtained from 22 persons with a stroke wearing a wearable sensor on the wrist, using statistical methods to infer tendencies of natural movements in daily activities. Here, we focused on three categories of statistical quantification: 1) those relating to distributions of acceleration and angular

velocity magnitudes of the forearm movement throughout the day; 2) those relating to the distribution of forearm postures experienced throughout the day; and 3) those related to the complexity of forearm movements performed throughout the day.

### *3.3. Methods*

#### *3.3.1. Wearable Sensor and Experimental Protocol*

The Manumeter is a wrist-worn device consisting of a six degrees-of-freedom (DOF) IMU with an accelerometer and a gyroscope (LSM6DSL; STMicrosystems, Switzerland), four magnetometers on four corners of the device, and an OLED display [40], [41]. For this study, we analyzed sensor signals from the IMU, sampled at 52.6 Hz obtained from a previous pilot study of the effectiveness of hand count feedback versus conventional home exercise [70]. Twenty-two participants (see overview of participants in Table I) wore the Manumeter once before the three-week intervention and then once after the intervention. 10 data sets were lost because of data acquisition problems [48], [77]. Thus, in total, 34 data sets were used for the study. The level of impairment of these participants was quantified using two common clinical measures. The Box and Blocks Test (BBT) requires participants to pick up small blocks from a box and transfer them over a divider, transferring as many blocks as possible in one minute [117]. The Upper Extremity Fugl Meyer (UEFM) test measures the ability of participants to perform 33 different test movements, rating each 0, 1, or 2 and summing the points to get a total possible score of 66 [118].

*Table 2. Characteristics of the 22 participants*

| | |
|---|---|
| Age | $57 \pm 15$ |
| Gender (Male[M])/Female([F]) | 16 M / 4 F |
| Time since stroke (months) | $40 \pm 33$ |
| Side of hemiparesis (Right [R]/Left [L]) | 12 R / 10 L |
| Type of stroke (Ischemic [I]/Hemorrhagic [H]) | 12 I / 10 H |
| Box and Blocks Test (Number of blocks transferred in 60 seconds) | $21 \pm 18$ |
| Upper Extremity Fugl-Meyer (UEFM) Score (0-66) | $40 \pm 13$ |

*3.3.2. Processing of the IMU Data*

We used the Madgwick filter to subtract the gravity components from acceleration [78]. To identify the statistical properties of the participant's arm activity, we needed to filter out periods of arm inactivity, which was done in the following way (Figure 9). First, we removed any time periods when sensor values remained constant for over three minutes (cf. [119], [120]). Second, we introduced two measures of arm activity: (1) the instantaneous upper limb use score $u(t)$ , and (2) the mean arm use score $\mathcal{U}(t)$. $u(t)$ is a binary number where 1 represents that the magnitude of the arm acceleration is over a small threshold, chosen as described below.

$$u_i(t) \triangleq \begin{cases} 0, & \text{UL is not in use at time } t \\ 1, & \text{UL is in use at time } t \end{cases} \quad (5)$$

$\mathcal{U}(t)$ is the average of u(t) over a sliding window with D = 10 seconds long:

$$\mathcal{U}_i(t;D) \triangleq \frac{1}{D}\int_{t-D}^{t} u_i(x)dx,\ t \in [D,T] \tag{6}$$

We identified a threshold for $\mathcal{U}(t)$ such that if $\mathcal{U}(t)$ were below that threshold, we considered the arm to be inactive at that movement and removed that data from our analysis. We will use the terms "total inactive time" for the total amount of time that we removed, and "total active time" for the time periods we kept. To choose thresholds for $u(t)$ and $\mathcal{U}(t)$, we conducted a grid search in the range of [0.05, 0.30] for the threshold of $u(t)$, and [0.05, 0.30] for the threshold of $\mathcal{U}(t)$, checking a combination of parameters that achieved a statistically significant correlation (see Supplementary Material). Our goal was to make the thresholds as lenient as possible to retain as much data as possible for analysis. However, we imposed the constraint that the chosen thresholds should produce a total inactive time that was as strongly correlated as possible with the impairment level measured by UEFM score. As a result, we chose a threshold of $u(t)$ = 0.1 G and a threshold of $\mathcal{U}(t)$ = 10 % as parameters for the rest of the analysis.

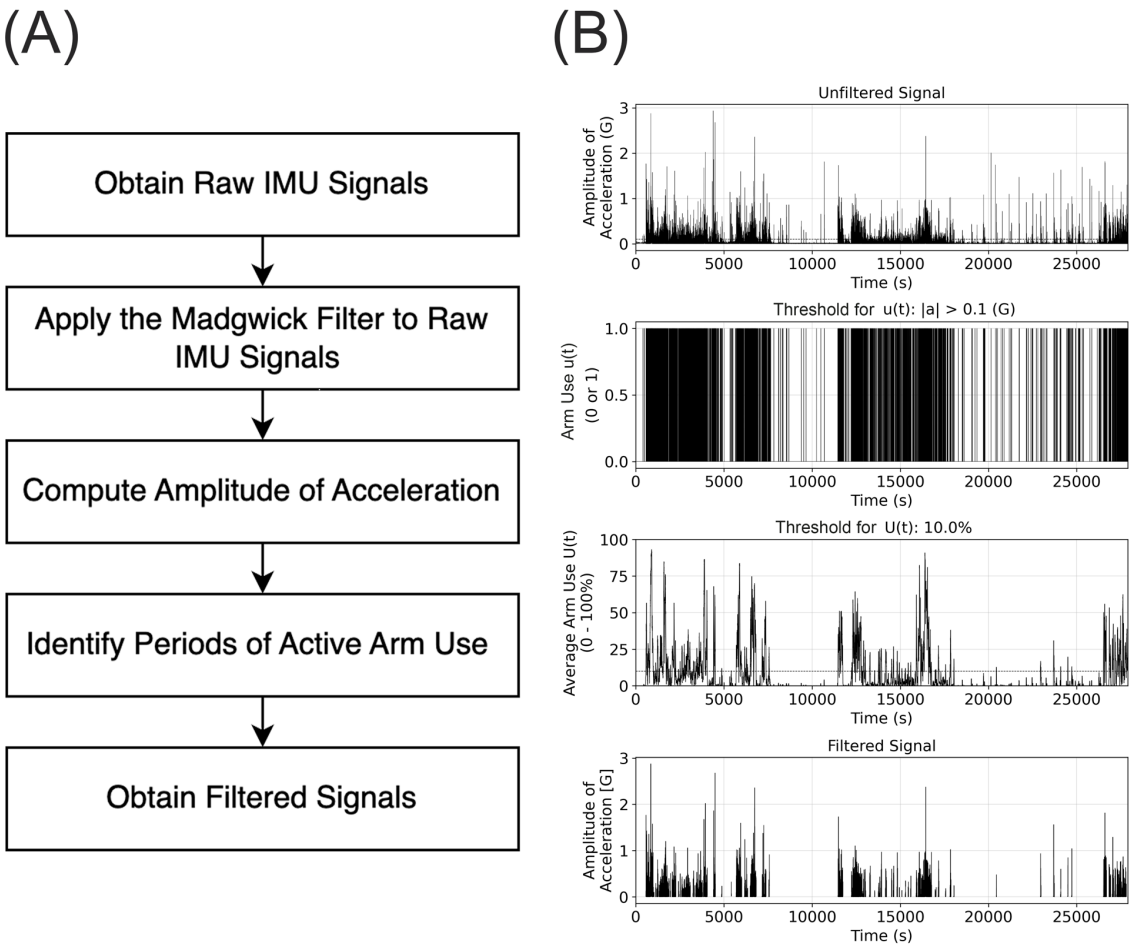



*Figure 9. Preprocessing of data to find active periods. (A) Overview of the preprocessing steps. The gravity component was subtracted using the Madgwick Filter and the periods of active arm use were identified. (B) Example of signals for one participant at different stages of preprocessing. From top to bottom: Raw amplitude of acceleration; Periods of active arm movement identified by the first threshold; Moving averages of filtered movements with a window size of 1000; Filtered acceleration using the second threshold.*

### *3.3.3. Forearm Orientation*

Leuenenberger et al. previously used an estimate of the orientation of an IMU worn on the wrist to quantify functionally relevant arm movement of stroke patients, proposing it as

a measure of movement quality [121]. Motivated by their approach and the principles in [122], [123], we similarly estimated device orientation using the law of cosines [122]:

$$\Theta = \arccos\left(\frac{-a_{ez}}{\sqrt{a_{ex}^2 + a_{ey}^2 + a_{ez}^2}}\right) \quad (7)$$

where $a_{ex}$, $a_{ey}$, and $a_{ez}$ represent the components of the measured acceleration vector along the x, y, and z axes, respectively, with respect to a sensor coordinate frame S, before Madgwick filtering (Figure 10). We normalize this vector to obtain a unit vector with a magnitude of 1. Equation 7 computes the angle Θ between the projection of the normalized acceleration vector in the sensor coordinate frame S, and a normal vector (0,0,1) with respect to a world coordinate W.

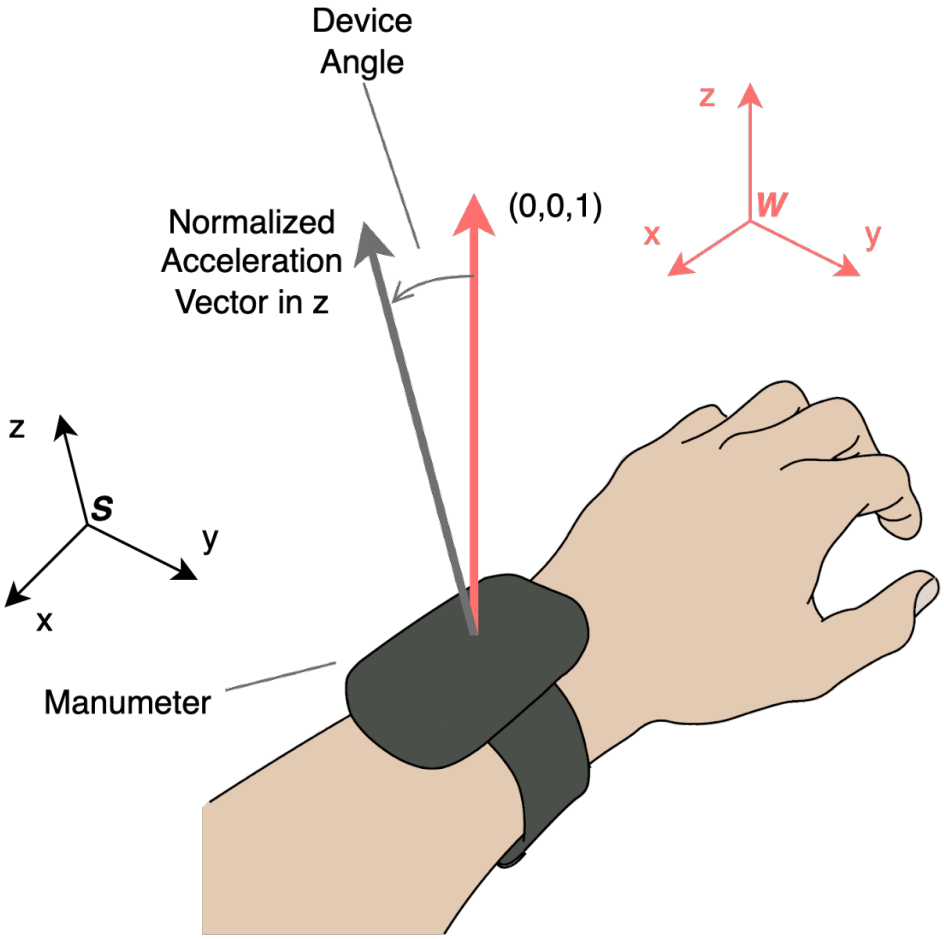



*Figure 10. The derivation of the device orientation with respect to acceleration vectors. The algorithm compares the tilt angle Θ between a vector in the world coordinate frame W and a vector in the sensor coordinate frame S.*

### *3.3.4. Sample Entropy*

To quantify the complexity of movement, we used Sample Entropy (SampEn), an established measure that quantifies the signal complexity of physiological measurements

[124]–[126], such as EEG signals [127] and EMG signals [128]. SampEn is defined as follows:

$$SampEn(m,r) = -\ln\left[\frac{B^{m+1}(\gamma)}{B^{m}(\gamma)}\right] = -\ln\left[\frac{(\mathrm{N}-\mathrm{m}-1)^{-1}\sum_{\mathrm{i}=1}^{\mathrm{N}-\mathrm{m}-1}\mathrm{B}_{\mathrm{i}}^{\mathrm{m}+1}(\gamma)}{(\mathrm{N}-\mathrm{m})^{-1}\sum_{\mathrm{i}=1}^{\mathrm{N}-m}\mathrm{B}_{\mathrm{i}}^{\mathrm{m}}(\gamma)}\right] \quad (8)$$

where $B^{m+1}$ represents the number of matches of length m+1 with i+1 th template, and $B^{m}$ represents the number of matches of length m with i th template. We need to choose two key parameters to assess movements: (1) a template length m, and (2) a tolerance r. m represents a length of template to compare the signal to the rest of the data. SampEn computes the number of matchings with a template having a length m (i.e., $B^{m}(\gamma)$), and m+1 (i.e., $B^{m+1}(\gamma)$), and then computes the ratio of matching counts between $B^{m+1}(\gamma)$ and $B^{m+1}(\gamma)$. The tolerance r checks if a difference between a template and an inspected window is acceptable. In addition, we optimized the following additional parameters: (3) the segmentation length N, (4) the sampling rate, and (5) the type of sensor signals. The sampling rate is important because SampEn may not be good at identifying a complexity of high-resolution sensor measurements due to the nature of the algorithm [129]. To assess quality of movement with a smaller sampling rate than an original rate (i.e., 52.6 Hz), we down sampled IMU signals from 52.6 Hz to 26, and 13 Hz, respectively. We applied SampEn to (i) the amplitude of acceleration, (ii) the amplitude of angular velocity, and (iii) the tilt angle computed by Equation 7.

To decide a best combination of parameters for (1)-(5) that maximizes a correlation between the sample entropy and UEFM scores, we chose parameters for each variable we studied (see Supplementary Material).

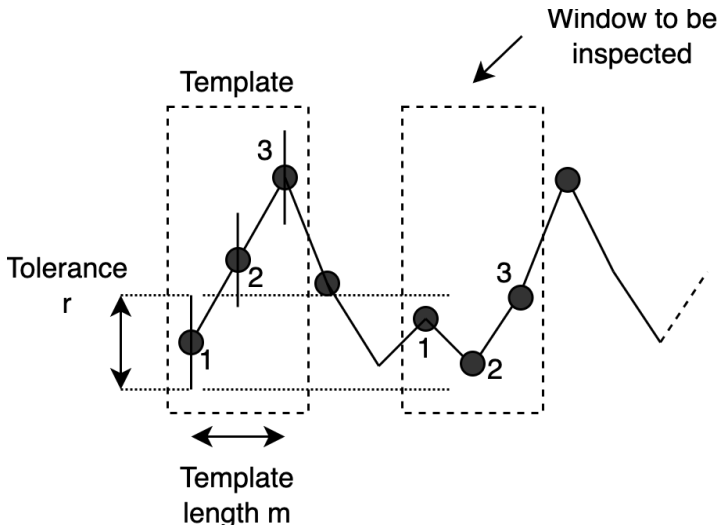



*Figure 11. Overview of the Sample Entropy (SampEn) calculation. SampEn checks if there is a similar sequence in the sliding window compared to the template window.*

### *3.3.5. Identifying Measure Dependence on Arm Impairment*

We hypothesized that subjects have a different statistical pattern of daily movement depending on their level of arm impairment and analyzed this dependence in two ways. First, we examined whether putative statistical measures of arm movement varied with UEFM score using regression analysis. Second, we compared groups of subjects divided into three discrete levels of impairment. A previous study [130] used Rasch analysis to propose that subjects could be split into severe, moderate, and mild groups based on UEFM cutoff scores of 19 and 47 out of 66. Here, we used a variant of this three-group approach based on our own analysis of clinical and wearable sensing data that we previously acquired, which allowed us to assign a distinct meaning for each group. Group 1 (severely impaired) and Group 2 (moderately impaired) are distinguished by their inability or ability to use their hand. Figure 12A shows the relationship between UEFM score and BBT score, a measure of hand function. We can see that the group of subjects with UEFM

score < 30 pts only scored 0-2 blocks on the BBT test, meaning they had little to no hand function. Groups 2 (moderately impaired) and 3 (mildly impaired) are distinguished by whether they use their hand in daily life. Figure 12B shows the relationship between hand use, measured with the Manumeter, and UEFM score, replotted from [48]. Subjects with UEFM score ≥ 50 points (approximately) show a greater amount of hand use in the real world. Thus, we used UEFM scores of 30 and 50 as thresholds to split subjects into three groups with distinct meanings, for whom we then compared the statistical properties of arm movement.

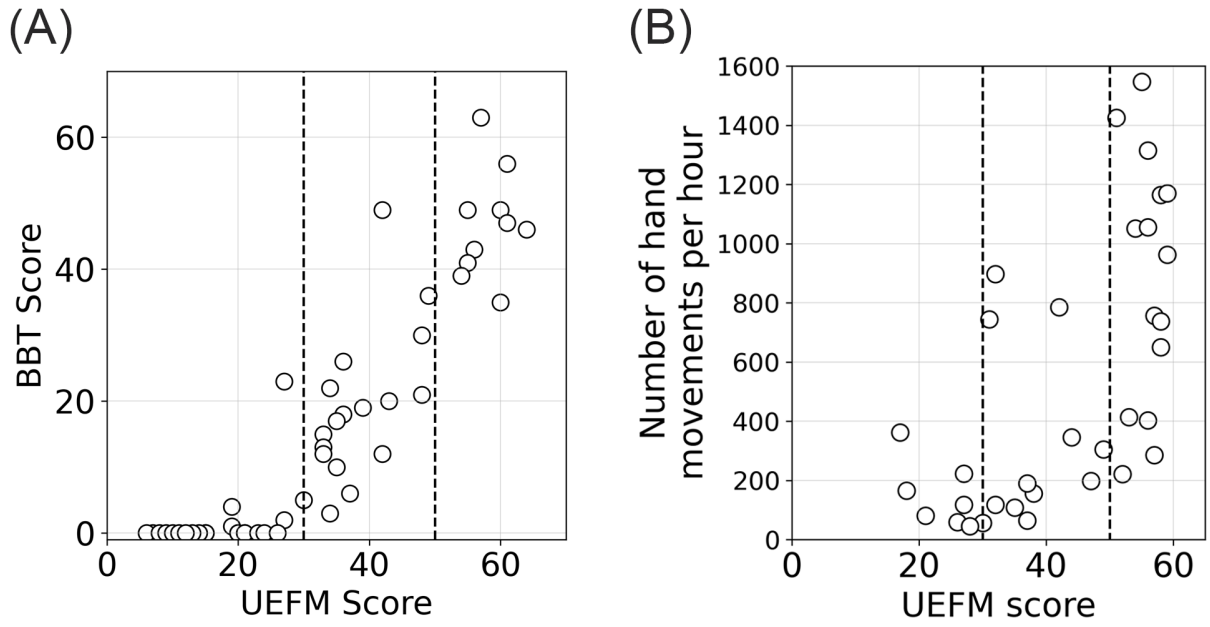



*Figure 12. Basis for selecting severe, moderate, mild impairment groups based on UEFM score thresholds (defined by the vertical dashed lines) (A) The relationship between BBT Score and UEFM Score, replotted from [131]. Hand function emerges around UEFM = 30 (first dashed line) (B) The relationship between hand use intensity and the UEFM score (replotted from [70]). The hand use intensity is the number of hand counts detected from the Manumeter. Daily hand use emerges around UEFM = 50.*

*3.3.6. Statistical analysis of movement experience quality features*

The left column of Table 1 provides a list of sensor variables (or "movement experience quality features") that we created, based on distributions of acceleration or speed, parameters describing the distribution of forearm postures, and signal complexity measures. We performed linear regression between each metric and the UEFM score. To compare statistical differences between the three groups assigned by impairment level, we used analysis of variance (ANOVA). In order to find a feature that distinguished one group from another, we conducted an unpaired t-test. To assess how much a group was different from another by a selected measure, we computed an effect size using Cohen's D [132].

We also used a multiple linear regression model to predict UEFM score. To check the multi collinearity between candidate features, we computed the variance inflation factor (VIF) between the variables:

$$VIF = \frac{1}{1 - R_i^2} \tag{9}$$

where $R_i^2$ represents the coefficient of determination. We conducted variable selection using the backward elimination [133]. In the ordinary least square (OLS) regression, we removed the variable having a highest p-value one by one until the model has a single variable. Then, we chose the model having a minimum Akaike Information Criterion (AIC), a measure assessing the quality of the model [134], [135]. AIC is often favored over $R^2$ because it takes into account both the goodness of fit and the complexity of the model, helping to avoid overfitting and select a more parsimonious model that is likely to

generalize better to new data [136], [137]. By minimizing AIC, moreover, we can select a model that provides a good balance between fitting the data well and having a lower complexity [138]. To compare the models, we computed differences between each AIC and the minimum AIC in backward elimination:

$$\mathrm{AIC} = 2\mathrm{k} - 2\ln\left(\hat{\mathrm{L}}\right) \tag{10}$$

$$\Delta = \mathrm{AIC} - \mathrm{AIC}_{\min} \tag{11}$$

where k represents a number of estimated parameters in the model, $\hat{\mathrm{L}}$ represents the maximum value of the loglikelihood function of the model, and $\mathrm{AIC}_{\min}$ denote the minimum AIC from a model. The minimum AIC is the value of AIC that corresponds to the best model among a set of candidate models. The difference between the minimum AIC and each AIC determines how acceptable it is to select a model; it is expected there is too much difference if AIC is larger than 2.0; Burnham and Anderson stated that there is a substantial empirical support if the delta of AIC is small enough (i.e., $\Delta < 2.0$) [139].

*Table 3. Summary of movement experience quality features calculated across the recorded movement experiences.*

| Kind | Sub kind | Correlation | $R^2$ | Effect size (p-value in t-test) | | | | ANOVA (Group 1,2,3) | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | Group 1 – Group 2 | Group 2 – Group 3 | Group 1 – Group 2,3 | Group 1,2 – Group 3 | f | p-value |
| Tilt Angle | Kurtosis | 0.383 (p=0.025) | 0.147 | 1.145(0.028) | 0.199(0.611) | 1.529(0.054) | 0.605(0.068) | 6.399 | 0.005 |
| | Skewness | 0.314 (p=0.07) | 0.099 | 0.292(0.548) | 0.563(0.158) | 0.616(0.156) | 0.663(0.064) | 2.049 | 0.146 |
| | Standard Deviation | 0.389 (p=0.023) | 0.151 | 0.438(0.369) | 0.395(0.348) | 0.734(0.093) | 0.592(0.096) | 2.049 | 0.146 |
| Ratio | Acceleration | 0.435 (p=0.01) | 0.189 | 0.561(0.255) | 0.830(0.042) | 0.791(0.072) | 1.037(0.005) | 4.699 | 0.016 |
| | Angular Velocity | 0.171 (p=0.333) | 0.029 | 0.841(0.095) | 0.445(0.261) | 0.644(0.139) | 0.079(0.821) | 1.69 | 0.201 |
| Mean | Acceleration | 0.341 (p=0.049) | 0.116 | 0.664(0.180) | 0.553(0.166) | 0.778(0.076) | 0.772(0.032) | 2.964 | 0.066 |
| | Angular Velocity | 0.188 (p=0.286) | 0.035 | 0.116(0.810) | 0.401(0.310) | 0.340(0.429) | 0.471(0.182) | 0.926 | 0.407 |
| Samp En | Acceleration | 0.405 (p=0.017) | 0.164 | 0.119(0.806) | 0.805(0.048) | 0.525(0.225) | 0.862(0.018) | 3.041 | 0.062 |
| | Angular Velocity | 0.41 (p=0.016) | 0.168 | 0.610(0.217) | 0.611(0.127) | 0.868(0.049) | 0.835(0.021) | 3.579 | 0.04 |
| | Tilt Angle | 0.5 (p=0.003) | 0.25 | 0.047(0.922) | 0.993(0.017) | 0.576(0.184) | 1.106(0.006) | 4.971 | 0.013 |

### *3.4. Results*

#### *3.4.1. Total active time*

The 22 participants with a stroke donned the wrist-worn IMU after they left the laboratory following several hours of clinical evaluations (i.e. usually in the late morning or afternoon). The average duration of IMU recording for the rest of the day was about six hours. During this time, participants performed active movements for 1.8 hours on average, as detected by the activity filter described in the Methods section. Total inactive time computed with respect to wear time for the impaired arm decreased significantly from about 80% to 60% as a function of the UEFM score (See figure in Supplementary Material).

*3.4.2. Distributions of acceleration and angular velocity magnitudes*

The distributions of the magnitude of acceleration and angular velocity differed for the groups of individuals in the three levels of impairment (Figure 13). Although the distributions were similar between groups, there were visible differences between groups for the lowest acceleration range [0-1 m/s$^2$], with the difference reversing in direction for the next highest acceleration range [1-2 m/s$^2$]. Specifically, the more impaired subjects spent more time at low accelerations, and less time at higher accelerations. To determine how strongly these differences related to UEFM score, we computed the ratio of the number of observations in the ranges, using a similar approach for angular velocity magnitude as well. There was a statistically significant correlation between the acceleration ratio and UEFM score (Figure 14A), but not the angular velocity ratio (Figure 14B). Less impaired subjects displayed a higher acceleration ratio, meaning they spent relatively more time at a higher acceleration magnitude range.

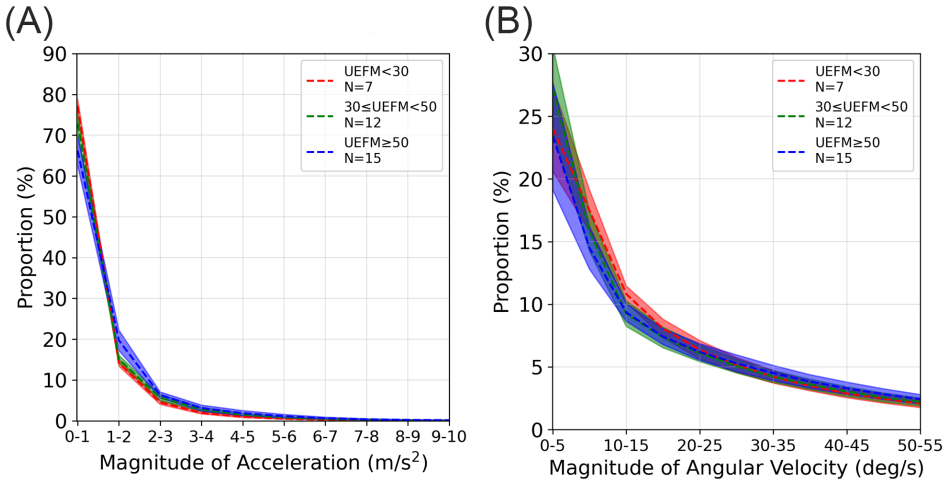



*Figure 13. The proportion of participants' movements as a function of magnitude of A) acceleration magnitude and B) angular velocity*

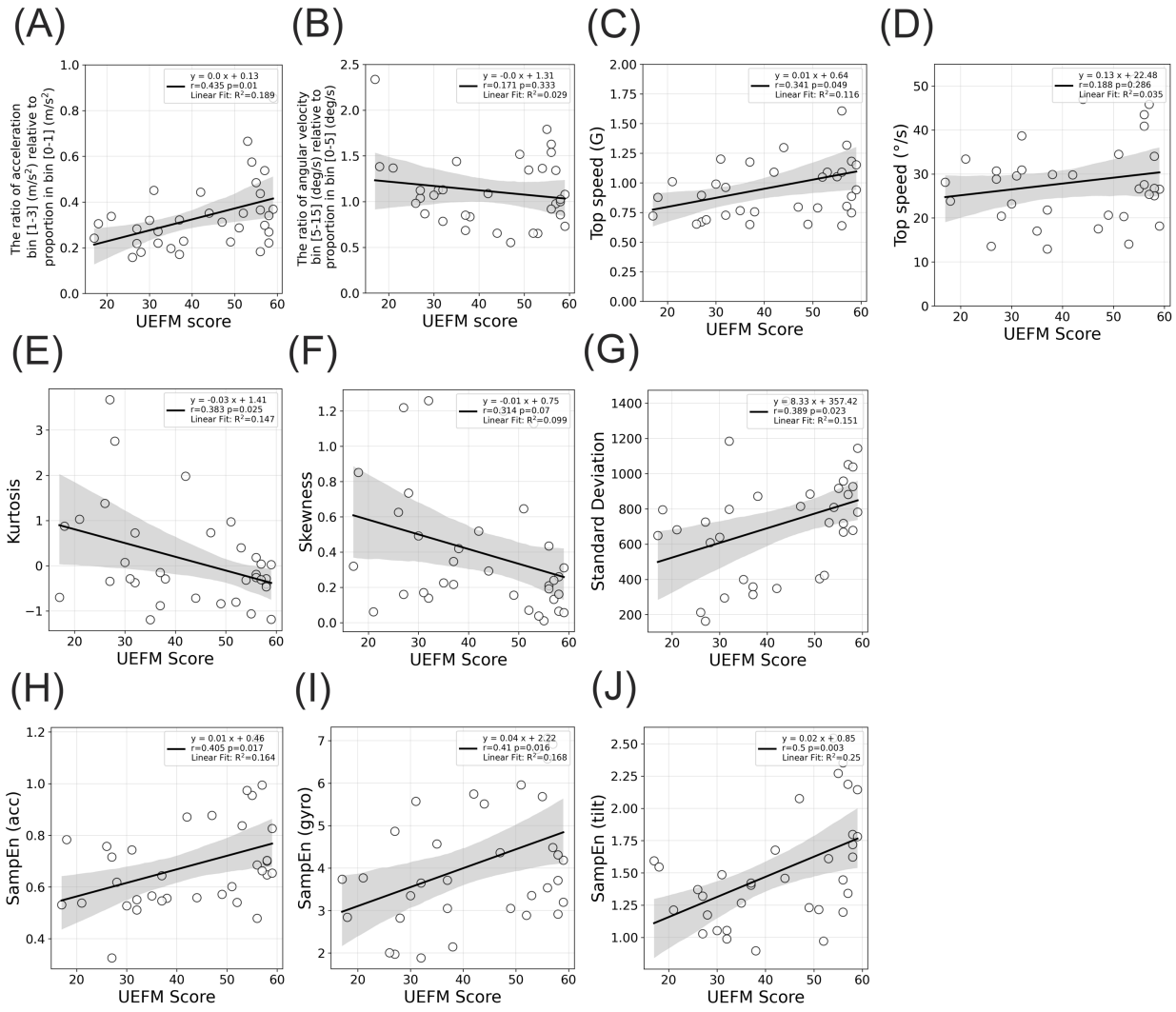



*Figure 14. Correlation between UEFM score and various sensor variables. The shaded regions represent the 95% confidence interval. The correlation between the UEFM score and: (A) the ratio of acceleration in [1-3] m/s$^2$ relative to [0-1] m/s$^2$, (B) The ratio of angular velocity in [5-15] deg/s relative to [0-5] deg/s. (C) the mean of acceleration, (D) the mean of angular velocity and the UEFM score, (E) Kurtosis of tilt angle, (F) skewness of tilt angle, (G) standard deviation of tilt angle, (H) SampEn of acceleration (I) SampEn of angular velocity, (J) SampEn of tilt angle.*

*3.4.3. Forearm posture with respect to gravity*

The distribution of forearm postures varied depending on group, with the severely impaired group showing a narrower distribution (Figure 15). To compare the shape of the statistical distribution of forearm posture, we used three standard statistical measures: kurtosis, skewness, and standard deviation. Kurtosis relates to the sharpness of the distribution; the distribution becomes more rounded as kurtosis increases [140]. Kurtosis decreased significantly as a function of UEFM score (Figure 14E). Skewness also decreased as a function of UEFM score, but this decrease only approached significance (Figure 14F). Finally, standard deviation increased significantly as a function of UEFM score (Figure 14G).

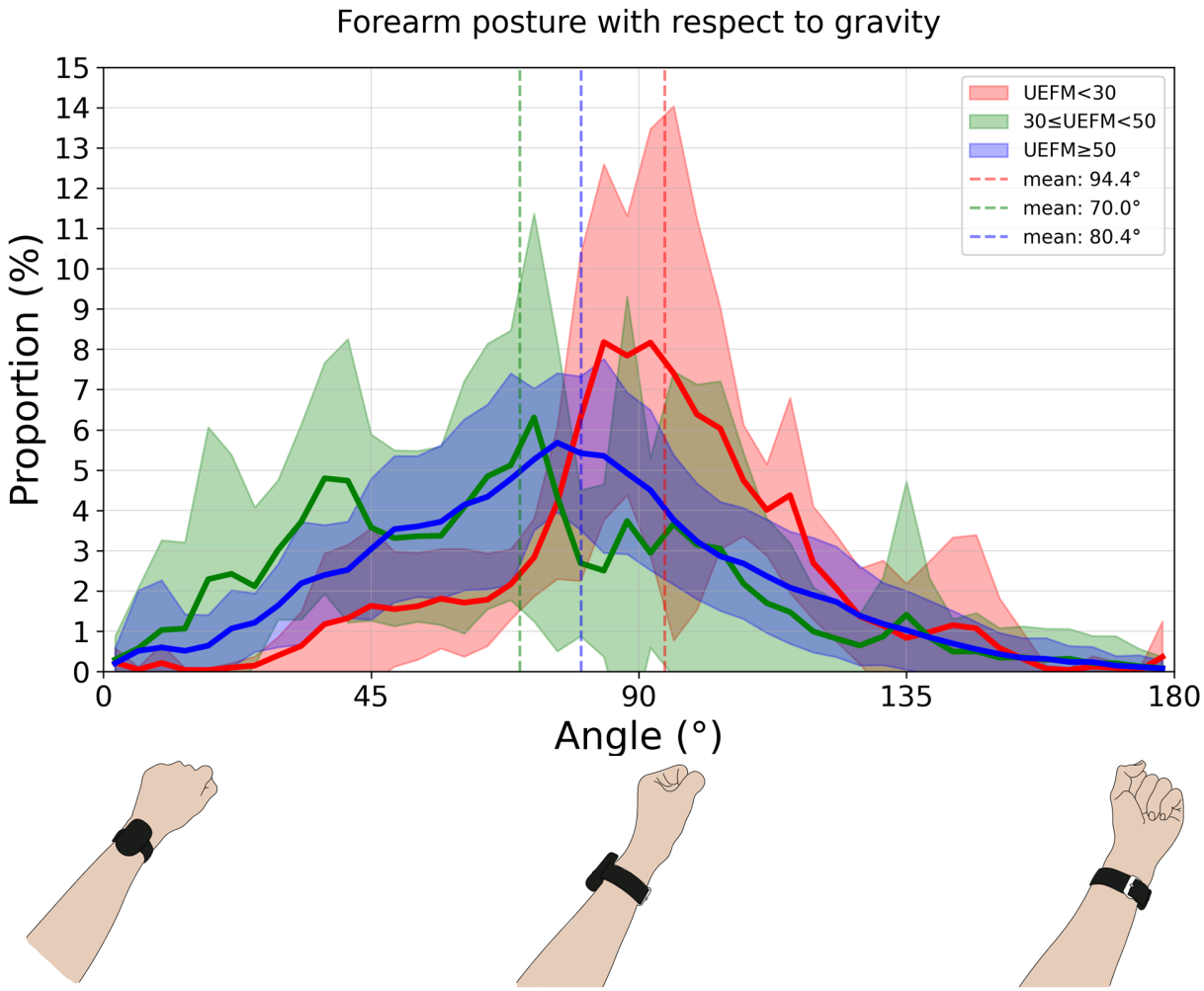



*Figure 15. The distribution of the device tilt angle and corresponding forearm postures with respect to the gravity, using 91 bins in range of 0-180 degrees (i.e., 2 degrees for one bin). The below three figures (B-D) correspond to the device orientation. The red, green, and blue lines represent the mean of distribution for Group 1, Group 2, and Group 3, respectively. The shaded areas represent the confidence interval for each group. The proportions represent the probability density distribution, such that the integration from 0-180 degrees is 100%.*

### *3.4.4. Sample entropy*

We also tested a measure of movement complexity – the sample entropy – applied to the acceleration magnitude, the angular velocity magnitude, and the estimated forearm

posture. For each of these measures, movement complexity increased significantly as a function of UEFM score (Figure 14H, I, and J).

*3.4.5. Identifying the measures that most effectively distinguished impairment level*

Table 3 summarizes which of the studied measures most effectively distinguished impairment level when subjects were grouped into the severe, moderate and mild impairment groups. Briefly, as a reminder, the group definitions were: Group 1: UEFM < 30, no hand function; Group 2: 30 ≤ UEFM < 50, hand function but low hand use; Group 3: UEFM ≥ 50, regular hand use. To distinguish Group 1 and Group 2, it was most useful to use kurtosis of the tilt angle distribution (Cohen's D = 1.1, t-test, p=0.028). To distinguish Group 2 and Group 3, it was most useful to use SampEn from tilt angle (Cohen's D = 0.99, t-test, p=0.017).

*3.4.6. Variable selection for the multivariable model that predicts the UEFM Score*

We sought to develop a multivariate linear model that predicted UEFM score. We first removed variables that were not statistically significantly correlated with the UEFM score (Ratio of Angular Velocity, p=0.333; Mean of Angular Velocity, p=0.286). Then we conducted a backward elimination process with the remaining variables (as described in detail in the Methods) to understand which combination of features best distinguished groups, removing variables from the mode l one by one (Table 4). We observed the lowest AIC when we selected the number of variables to be two (Figure 16A). In this case, the selected variables were the kurtosis of the tilt distribution (i.e. forearm postural diversity) and the sample entropy of tilt (i.e., forearm postural complexity). We further examined the

correlation coefficients between the variables (Figure 8B), seeking to identify if there was multicollinearity between any variables, which is a statistical phenomenon that occurs when two or more independent variables in a multiple regression model are highly correlated; a high correlation makes it difficult to determine the individual contribution of each variable to the dependent variable. To assess multicollinearity, we used the criterion that the Variance Inflation Factor (VIF) should be less than 5 [141]. The VIF measures the extent to which the variance of the estimated regression coefficients is increased due to multicollinearity. A VIF value less than 5 is considered acceptable, indicating that multicollinearity is not a significant concern. In our analysis, the correlation between Kurtosis and SampEn of tilt angle was weak (Pearson Correlation Coefficient= 0.22), and the VIF in Step 8 was 1.05 for both variables. These results suggest that multicollinearity is not a major issue for these variables in the model.

*Table 4. Steps of the backward eliminations*

| Step | Variable | Coefficient | Standard Error | T | P-value | [0.025 | 0.975] | Variance Inflation Factor (VIF) |
|---|---|---|---|---|---|---|---|---|
| 1 | Constant | 21.07 | 10.74 | 1.96 | 0.06 | -1.06 | 43.2 | 26.48 |
| | Kurtosis of Tilt Angle | -1.96 | 2.93 | -0.67 | 0.51 | -7.99 | 4.08 | 2.25 |
| | Skewness of Tilt Angle | -2.57 | 8.73 | -0.29 | 0.77 | -20.55 | 15.42 | 2.24 |
| | Standard Deviation of Tilt Angle | 0.01 | 0.01 | 0.84 | 0.41 | -0.01 | 0.03 | 1.37 |
| | Ratio of Acceleration | 24.53 | 20.26 | 1.21 | 0.24 | -17.19 | 66.25 | 2.08 |
| | Mean of Acceleration | -12.08 | 12.02 | -1 | 0.33 | -36.83 | 12.68 | 3.04 |
| | Sample Entropy (Acceleration) | -5.84 | 24.48 | -0.24 | 0.81 | -56.26 | 44.58 | 4.17 |
| | Sample Entropy (Angular Velocity) | 1.88 | 2.2 | 0.86 | 0.4 | -2.65 | 6.42 | 2.39 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Sample Entropy (Tilt) | 11.64 | 10.2 | 1.14 | 0.27 | -9.37 | 32.65 | 4.22 |
| 2 | Constant | 20.53 | 10.31 | 1.99 | 0.06 | -0.67 | 41.73 | 25.31 |
| | Kurtosis of Tilt Angle | -1.98 | 2.88 | -0.69 | 0.5 | -7.89 | 3.94 | 2.25 |
| | Skewness of Tilt Angle | -2.83 | 8.5 | -0.33 | 0.74 | -20.31 | 14.64 | 2.2 |
| | Standard Deviation of Tilt Angle | 0.01 | 0.01 | 0.86 | 0.4 | -0.01 | 0.02 | 1.37 |
| | Ratio of Acceleration | 24.24 | 19.85 | 1.22 | 0.23 | -16.56 | 65.04 | 2.07 |
| | Mean of Acceleration | -12.18 | 11.79 | -1.03 | 0.31 | -36.42 | 12.06 | 3.03 |
| | Sample Entropy (Angular Velocity) | 1.74 | 2.08 | 0.84 | 0.41 | -2.54 | 6.03 | 2.22 |
| | Sample Entropy (Tilt) | 9.91 | 7.03 | 1.41 | 0.17 | -4.55 | 24.37 | 2.08 |
| 3 | Constant | 19 | 9.07 | 2.09 | 0.05 | 0.39 | 37.61 | 20.25 |
| | Kurtosis of Tilt Angle | -2.61 | 2.13 | -1.23 | 0.23 | -6.98 | 1.76 | 1.27 |
| | Standard Deviation of Tilt Angle | 0.01 | 0.01 | 0.85 | 0.4 | -0.01 | 0.02 | 1.36 |
| | Ratio of Acceleration | 23.28 | 19.31 | 1.21 | 0.24 | -16.35 | 62.91 | 2.03 |
| | Mean of Acceleration | -11.34 | 11.33 | -1 | 0.33 | -34.59 | 11.91 | 2.89 |
| | Sample Entropy (Angular Velocity) | 1.87 | 2.01 | 0.93 | 0.36 | -2.26 | 6 | 2.14 |
| | Sample Entropy (Tilt) | 9.67 | 6.88 | 1.41 | 0.17 | -4.44 | 23.79 | 2.06 |
| 4 | Constant | 22.25 | 8.18 | 2.72 | 0.01 | 5.51 | 39 | 16.61 |
| | Kurtosis of Tilt Angle | -3.31 | 1.95 | -1.7 | 0.1 | -7.31 | 0.69 | 1.08 |
| | Ratio of Acceleration | 25.35 | 19.06 | 1.33 | 0.19 | -13.69 | 64.4 | 2 |
| | Mean of Acceleration | -11.05 | 11.27 | -0.98 | 0.34 | -34.13 | 12.03 | 2.89 |
| | Sample Entropy (Angular Velocity) | 1.83 | 2 | 0.92 | 0.37 | -2.27 | 5.94 | 2.14 |
| | Sample Entropy (Tilt) | 10.29 | 6.81 | 1.51 | 0.14 | -3.66 | 24.23 | 2.04 |
| 5 | Constant | 22.72 | 8.14 | 2.79 | 0.01 | 6.08 | 39.36 | 16.55 |
| | Kurtosis of Tilt Angle | -3.47 | 1.94 | -1.79 | 0.08 | -7.43 | 0.5 | 1.07 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Ratio of Acceleration | 22.65 | 18.78 | 1.21 | 0.24 | -15.76 | 61.06 | 1.95 |
| | Mean of Acceleration | -6.2 | 9.92 | -0.63 | 0.54 | -26.5 | 14.09 | 2.25 |
| | Sample Entropy (Tilt) | 12.48 | 6.36 | 1.96 | 0.06 | -0.52 | 25.48 | 1.79 |
| 6 | Constant | 21.15 | 7.66 | 2.76 | 0.01 | 5.5 | 36.79 | 14.97 |
| | Kurtosis of Tilt Angle | -3.39 | 1.91 | -1.77 | 0.09 | -7.3 | 0.52 | 1.07 |
| | Ratio of Acceleration | 16.88 | 16.19 | 1.04 | 0.31 | -16.18 | 49.94 | 1.48 |
| | Sample Entropy (Tilt) | 10.82 | 5.71 | 1.89 | 0.07 | -0.85 | 22.49 | 1.47 |
| 7 | Constant | 22.04 | 7.62 | 2.89 | 0.01 | 6.49 | 37.59 | 14.79 |
| | Kurtosis of Tilt Angle | -3.64 | 1.9 | -1.91 | 0.07 | -7.52 | 0.24 | 1.05 |
| | Sample Entropy (Tilt) | 14.02 | 4.83 | 2.9 | 0.01 | 4.17 | 23.87 | 1.05 |
| 8 | Constant | 18.58 | 7.71 | 2.41 | 0.02 | 2.88 | 34.29 | 13.96 |
| | Sample Entropy (Tilt) | 16.03 | 4.91 | 3.27 | 0 | 6.04 | 26.02 | 1 |

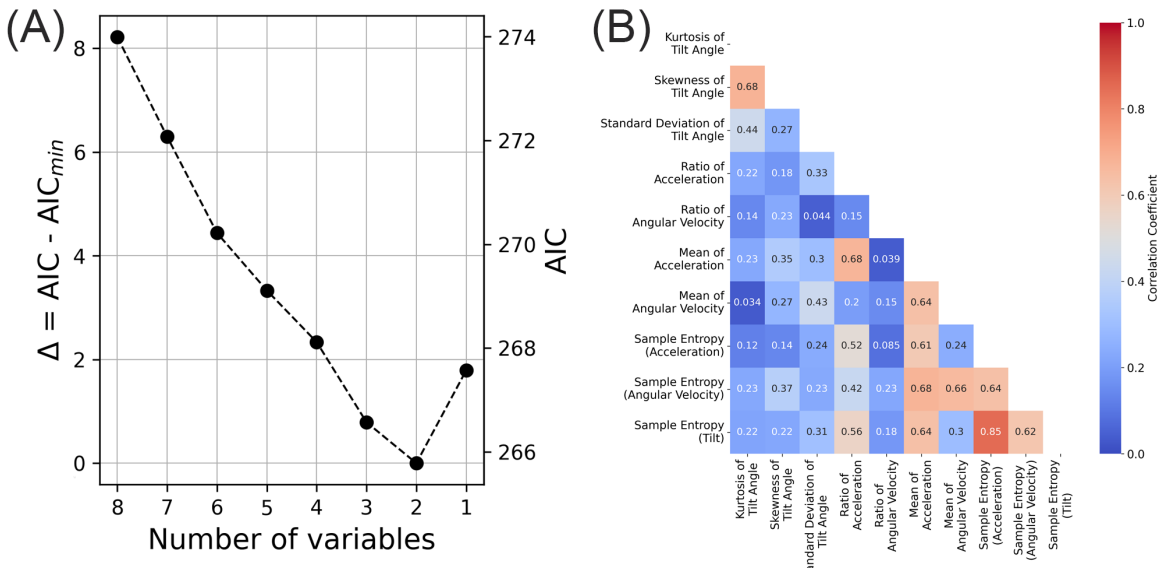



*Figure 16. The variable selection analysis. (A) Model comparison through the AIC. (B) Correlations between features.*

### *3.5. Discussion*

The objective of this study was to identify statistical characteristics of daily arm movements that became more prominent as arm impairment decreased, based on data obtained from a wrist IMU worn by 22 chronic stroke participants during their day. Our focus here was on the statistical distributions of movement features across many arm movements made during the period the wearable sensor was worn. We hypothesize these features relate to the "quality of the movement experience" (which we will abbreviate QOME for this Discussion) over a period time. We identified several QOME-related measures that increased as UE Fugl-Meyer (UEFM) score increased: forearm speed, forearm postural diversity (quantified by kurtosis of the tilt-angle), and forearm postural complexity (quantified by sample entropy of tilt angle). Dividing participants into severe, moderate, and mild impairment groups, we found that forearm postural diversity and complexity were best able to distinguish the groups. Specifically, to distinguish between severe and moderate impairment, kurtosis of the tilt angle, our measure of forearm postural diversity, was most effective. To distinguish between moderate and mild impairment, SampEn of the tilt angle, our measure of movement complexity, was most effective. The application of a multivariate modeling approach confirmed that these were also the best variables from among the ones we considered for predicting UEFM score. We discuss now these results as well as limitations and future directions.

#### *3.5.1. Quantifying movement diversity based on forearm posture (tilt angle) distribution*

We did not attempt to estimate the three angles that are needed to fully describe forearm posture over time, which is challenging with an IMU, but rather examined the

simpler idea of tilt angle analysis with respect to gravity direction, first proposed by Leuenberger et al. [121] in the context of wearables for stroke rehabilitation. They found that the impaired arm had lower tilt angle (or “elevation”) during performance of activities of daily living, found different probability distributions between the impaired arm and unimpaired arm, and suggested forearm posture as a potential target for improving patients’ upper arm dexterity. We found that tilt angle was distributed across a broad range of angles for participants in each impairment level, but more narrowly concentrated for the most severely impaired group. We found that the established statistical concept of kurtosis of a distribution was useful to quantify this narrowing.

A narrow tilt-angle concentration is consistent with the concept of “postural stagnation” – i.e. that persons with severe impairment after stroke tend to keep their arm in a stereotypical posture when not using it. Previous studies have identified five different arm postures that stroke patients adopt, with two being most prevalent [142], [143]. Further, postural stagnation in severe impairment is consistent with the observation of stereotypical, abnormal synergistic movement patterns due to reduced capability of the corticospinal tract and/or the presence of spasticity [116]. A study of the ability of people with stroke to reach in a wide range of directions found that severely impaired individuals were constrained to reach in a narrow range [144]. Based on these results and the experimental findings, it seems probable that postural stagnation is a valuable concept for QOME metrics. As mentioned in the Introduction, if one only ever practices the same thing (i.e. holds the arm in a limited set of postures), it is difficult to improve in skill (i.e. learn to activate muscles for a wide range of arm postures and activities).

Limitations of the tilt angle approach are that it cannot distinguish forearm supination/ pronation from shoulder internal/external rotation, and it doesn't consider the posture of the whole body. Thus periods spent lying or reclining might introduce noise into the distribution. Nevertheless, we found tilt-angle to be a valuable feature for distinguishing severely and moderately impaired participants.

In the context of motor disorders, there is often an imbalance in movement patterns. On one hand, you have conditions like stroke-induced paralysis, which often result in reduced movement complexity and diversity due to partial or complete loss of motor function. On the other hand, conditions like Parkinson's Disease (PD) and dystonia are often characterized by excessive, uncontrolled, and complex movements. For instance, patients with PD might experience tremors or other involuntary movements, while those with dystonia might have abnormal, often complex, postures due to sustained muscle contractions. In these cases, the goal is often to reduce the complexity and diversity of movements, aiming to gain more control and move back towards the "normal" spectrum. Thus, this study may suggest the goal of rehabilitation in these contexts could be seen as calibrating the motor control functions of the nervous system to maintain operation within an optimal range.

### *3.5.2. Quantifying movement complexity*

To quantify movement complexity we focused on one of the many possible measure of signal complexity – entropy, a measure that has found application in human movement science for analyzing postural control, walking activity, spontaneous leg activity in infants, and finger force production (see review: [124]). Entropy features, specifically

sample entropy, have been increasingly utilized in studies with wearable sensors, analyzing various biological signals to improve detection and prediction of motor impairments. In Parkinson's disease research, Shawen et al. applied sample entropy to acceleration and angular velocity signals from skin-mounted sensors and smartwatches, enhancing tremor detection accuracy [145]. In autism studies, Konrad et al. used sample entropy to analyze wrist accelerometer acceleration signals, finding a moderate relation to motor coordination [146]. For post-stroke rehabilitation, O'Brien et al. applied sample entropy to both acceleration and rotational velocity signals from inertial measurement unit (IMU) sensors, linking higher sample entropy values to higher ambulation levels [147]. After stroke, sample entropy has been used to gain insight into UE muscle activity changes following robotic rehabilitation [148]. To our knowledge, this is the first report of showing the potential value of sample entropy of forearm posture, measured from a wristworn IMU, for gaining insight into UE impairment after stroke. SampEn best distinguished participants with moderate and mild impairment, with the more mildly impaired group showing more complexity in their movements. This seems indicative that they typically achieved a richer daily movement experience, perhaps because they could use the arm in a wider variety of activities.

A limitation of the use of sample entropy is that it is strongly dependent on parameter selection [124]. To address this, we used a grid search to find the best combination of parameters (see Supplementary Material). Another potential limitation, particularly with a view toward implementation in a wearable sensor, is that SampEn can be computationally costly, with a complexity of $O\left(N^{2-\frac{1}{m+1}}\right)$, where N is the segmentation length and m is the

template length [125]. Nevertheless, at least one study was able to implement SampEn in a wearable sensor for sleep research [149].

### *3.5.3. Other considerations: movement activity and speed*

Based on the inactivity filtering results, we found that participants with more severe impairment exhibited reduced UE activity. This finding is consistent with the theory of learned non-use [106] as well as previous studies using wearable sensing (e.g. [48], [150]). It was unexpected to us that movement diversity (quantified by kurtosis of tilt angle) and complexity (quantified by sample entropy) were the best features for distinguishing groups and predicting UEFM score, as opposed to movement speed. Nevertheless, distributions of movement speed did have some power to discriminate impairment level. In a study relevant for considering the potential of QOME wearable feedback, Dejong et al. examined the effect of instructing individuals with a stroke to move the UE more quickly as they reached and grasped a cup. They found that not only could the participants move their upper arm more quickly, but also that movement quality improved, as assessed by straighter reach paths and larger hand grip apertures [151]. Thus, by focusing on an easily instructed variable, other benefits related to movement quality could be obtained.

### *3.5.4. Limitations and Future Work*

Besides the specific limitations mentioned above with respect to the diversity and complexity measures, there are several other limitations to this work. First, we studied a relatively small list of potential measures related to QOME. Other measures are certainly

possible and may be more powerful predictors. Second, the motivation for this work was to identify potential QOME measures that could be provided by a wrist-worn sensor to enhance rehabilitation. However, we have not shown there is a causal relationship between providing QOME measures and recovery. It may not be useful for patients to practice a movement that increases QOME, although a large body of motor learning research does support the idea that challenging and variable task practice is beneficial [46]. Further, we have not identified whether and how people with a stroke can volitionally change a QOME measure. We aim to address these questions by embedding real-time QOME metrics in a wearable sensor and studying the effect of providing QOME feedback in future clinical trials.

# Chapter 4. Quantifying movement complexity and diversity during upper extremity activities

## *4.1. Summary of the Chapter*

Over 80 percent of people who have experienced a stroke incur upper extremity (UE) impairment resulting in reduced arm use in daily life. A few studies have examined the use of wearable feedback of the daily quantity of arm movement in order to promote recovery, but with limited success. We posit that it may be more effective to encourage an increase in beneficial movement experiences, rather than simply the overall amount of movement each day. In the last chapter, we identified two key measures that increased as UE Fugl-Meyer (UEFM) score increased: forearm postural diversity (quantified by kurtosis of the tilt-angle), and forearm postural complexity (quantified by sample entropy of tilt angle). Based on these findings, we posit that encouraging people to achieve more forearm postural diversity and complexity might be therapeutically beneficial. But what activities could one suggest helping patients achieve this goal? We recruited 7 unimpaired individuals and evaluated a set of 12 therapeutic activities for postural diversity and complexity. The activities were performed while seated and included: a set of conventional rehabilitation therapy exercises for the hand and arm, ping-pong, the card games Speed and Klondike, American Sign Language, Tai-chi, Cornhole, Nintendo Switch sports, balloon volleyball, cup stacking, and FitMi exercises, where FitMi is a commercial sensor system designed to guide upper extremity rehabilitation exercises. The participants performed each exercise for 10 minutes while wearing a wrist accelerometer and we computed sample entropy and kurtosis of the forearm tilt angle with sliding windows of 1,

3, and 5 minutes. Engaging in conventional rehabilitation therapy exercises created high values for forearm postural diversity but not complexity. Playing the card game Speed and exercising with the commercial sensor system produced the highest values for both postural diversity and complexity. Values for diversity and complexity were consistent across participants and data analysis window sizes of 1, 3, and 5 minutes. These results suggest that different candidate activities cause movement experiences with differing levels of movement diversity and complexity. Diversity and complexity can be accurately estimated with a 1minute window.

*4.2. Introduction*

The aftermath of a stroke can be a challenging time for patients, particularly for those who experience upper extremity (UE) impairment, resulting in limited ability to use the arm in daily life. Several studies have explored the potential of using wearable feedback to promote recovery by increasing daily arm movement [48], [70]. However, these studies have had limited success in substantially reducing in arm impairment as measured by clinical scales [70]. This leads us to hypothesize that it may be more beneficial to encourage beneficial patterns of movement rather than simply increasing overall arm movement.

In the last chapter, we presented forearm postural diversity (quantified by kurtosis of the tilt-angle), and forearm postural complexity (quantified by sample entropy of tilt angle) as measures that can be implemented on a wrist-worn Inertial Measurement Unit (IMU) to track whether the daily movement experienced by people with stroke is impoverished by their impairment. Then, we propose to use wearable sensing to increase

forearm postural diversity and complexity as a means of promoting post-stroke rehabilitation.

To achieve this goal, it would be desirable to be able to suggest beneficial activities to people with stroke that they could engage in to increase their experience of postural diversity and complexity. To that end, we evaluated a set of 12 therapeutic activities to identify which cause the highest values of postural diversity and complexity. We analyzed data obtained from wrist IMUs worn by 7 unimpaired patients as they performed the activities in the laboratory. We also explored the length of time needed to acquire enough data to sufficiently identify how each activity affects forearm postural diversity and complexity.

## *4.3. Methods*

### *4.3.1. Metrics for quantifying movement complexity and diversity*

In this study, we used two metrics to analyze the arm movement of the participants: sample entropy and kurtosis of forearm tilt angle (see Chapter 3). We measured tilt angle using the method introduced by [123], [152]. Sample entropy measures the predictability of a time series, with higher values indicating more irregular data [124], [125], [129]. Kurtosis measures the sharpness of the peak of a frequency-distribution curve, with higher values indicating a sharper peak [140].

*4.3.2. Participants and equipment*

Participants in this study were instructed to wear an IMU wrist-band (MiGo, Flint Rehabilitation Devices LLC, Irvine, CA, USA). The device contains a six-axis IMU with a sampling frequency of 200 Hz. The device streamed data at 100 Hz to a receiver (nrf52840, Nordic Semiconductor, Norway) connected to a PC. Seven unimpaired participants without any history of neurological disorders, musculoskeletal injuries or surgeries, or other conditions that could affect their upper limb movements, were recruited for this study. All participants were aged 20-25 and were male. The study was deemed exempt by the UCI Institutional Review Board and participants provided informed consent.

*4.3.3. Experimental protocol*

The participants were asked to perform 12 different activities while wearing the MiGo on their dominant hand. These activities were selected based on their potential to promote postural diversity and complexity. we included: Ping-pong, a table tennis game that requires hand-eye coordination and reflexes; Playing cards Speed, a fast-paced card game that requires users' reaction time and dexterity; Playing cards Klondike, a solitaire card game that requires strategy and concentration, and Speed, another fast-paced card game that tests players' reaction time and hand dexterity; American Sign Language, which involves learning and practicing sign language requiring fine motor skills and hand dexterity; Cornhole, a lawn game that involves throwing bean bags at a raised platform with a hole, requiring hand-eye coordination and targeting skills; Tai-chi, a form of exercise that promotes relaxation, balance, and coordination through slow, controlled movements; Nintendo Switch sports Chambara and Tennis, virtual reality video games that involve

physical movements and coordination, to engage participants in a fun and interactive way; FitMi exercises, a commercial sensor system for guiding upper extremity exercises after stroke, promoting rehabilitation and recovery [153]; Standard therapy exercises, which involve conventional rehabilitation therapy exercises for the hand and arm to regain strength, flexibility, and function; Balloon volleyball, a low-impact game that encourages hand-eye coordination and upper body movement by volleying a balloon back and forth; Cup stacking, a timed sport that challenges participants to stack and unstack cups in specific sequences.

Table 5 shows the number of participants who performed each activity.

*Table 5. Number of participants who performed each of the 12 activities*

| Task | Number of Participants |
|---|---|
| Ping-Pong | 7 |
| Playing Cards (Speed) | 6 |
| Playing Cards (Klondike) | 7 |
| Sign Language | 7 |
| Cornhole | 7 |
| Tai-Chi | 6 |
| Switch Sports (Chambara) | 7 |
| Switch Sports (Tennis) | 7 |
| FitMi | 6 |
| Standard Therapy | 6 |

| | |
|---|---|
| Balloon Volleyball | 7 |
| Cup Stacking | 7 |

Participants were given a brief introduction to each activity and allowed to practice for a few minutes to ensure their comfort with the task. They were then instructed to perform each activity for 10 minutes while seated except for ping-pong and cornhole. The MiGo was worn on the dominant side of the wrist throughout each exercise.

For ping-pong, a portable ping-pong court (FBSPORT, California, USA) was used, as it was easily adaptable to different environments. The motivation for choosing this portable ping-pong court was the recognition that most people do not have access to a full-size ping-pong table at home. By using a portable court, the study aimed to simulate a more realistic scenario, reflecting the typical conditions in which participants might engage in the activity.

For the conventional rehabilitation exercise activity, four exercises from a list of 20 provided to patients participating in a control group in a previous clinical trial of sensor-assisted therapy were selected for the standard therapy exercises provided to the clinical trial with another wrist-worn tracker, the Manumeter [70]. These exercises were designed by experience rehabilitation therapists and encompass various aspects of arm rehabilitation, targeting the shoulder, elbow, wrist, and fingers, with the overarching aim of improving range of motion, minimizing stiffness and pain, and providing sensory input to the weaker arm. The chosen four exercises were hypothesized to generate significant complexity and diversity in movement. The selected exercises included Weight Bearing From Sitting (Elbow/Wrist/Finger Extension), Arm Flip (Forearm Supination/Pronation),

Chopping Wood (Elbow Flexion/Extension), and Forearm Supination/Pronation at Tabletop.

For the exercise segment that used the sensorized rehabilitation exercise device (FitMi), participants performed a series of arm exercises using the device in a randomized order. Each exercise took approximately 1-2 minutes to complete, and the entire set of exercises lasted for a total of 10 minutes. We chose the following FitMi exercises as they involve arm movements; Reach to Target #1, Clapping, Reach to Target #2, Reach to Target #3, Wrist Supination, Bicep Curls, Shoulder Flexion, Shoulder Abduction, and Fly Out.

*4.3.4. Data Analysis*

Data from the MiGo were collected at a sampling rate of 100 Hz, then downsampled to 52.6 Hz using the scipy.resample function [76]. Data were segmented into 1-minute, 3-minute, and 5-minute windows, shifting by one second (i.g., 52 samples at a time). For each window, the tilt angle of the forearm was computed, and then sample entropy and kurtosis were calculated from the tilt angle (see previous chapter for parameter selection). The tilt angle was defined as the angle between the forearm and the horizontal plane (see the previous chapter). Due to data acquisition issues, one data set was lost for Switch Sports (Chambara) and Switch Sports (Tennis).

For every window, sample entropy and kurtosis of the tilt angle were calculated. We initially obtained sample entropy and kurtosis using an N-minute sliding window; for example, a 1-minute window corresponds to 3,120 samples (1 minute = 60 seconds x 52 Hz). Likewise, the 3- and 5-minute windows had 9,360 and 15,600 samples,

respectively.We shifted this window with 1 second stride (i.e., 52 data samples), and the mean values of sample entropy and kurtosis were computed for each activity and each window size. These mean values were then compared across different activities and window sizes to assess the potential of each activity in promoting postural diversity and complexity.

### *4.3.5. Results*

Figure 17 shows the mean of sample entropy and kurtosis of the tilt angle using sliding windows of 1, 3, and 5 minutes, with a 1-second stride (where stride is how far the window moves in every step) quantified from 12 exercises. Conventional rehabilitation therapy exercises generated high values for forearm postural diversity but did not significantly contribute to its complexity. The activities that yielded the highest values for both postural diversity and complexity were playing the card game Speed and exercising with the FitMi.

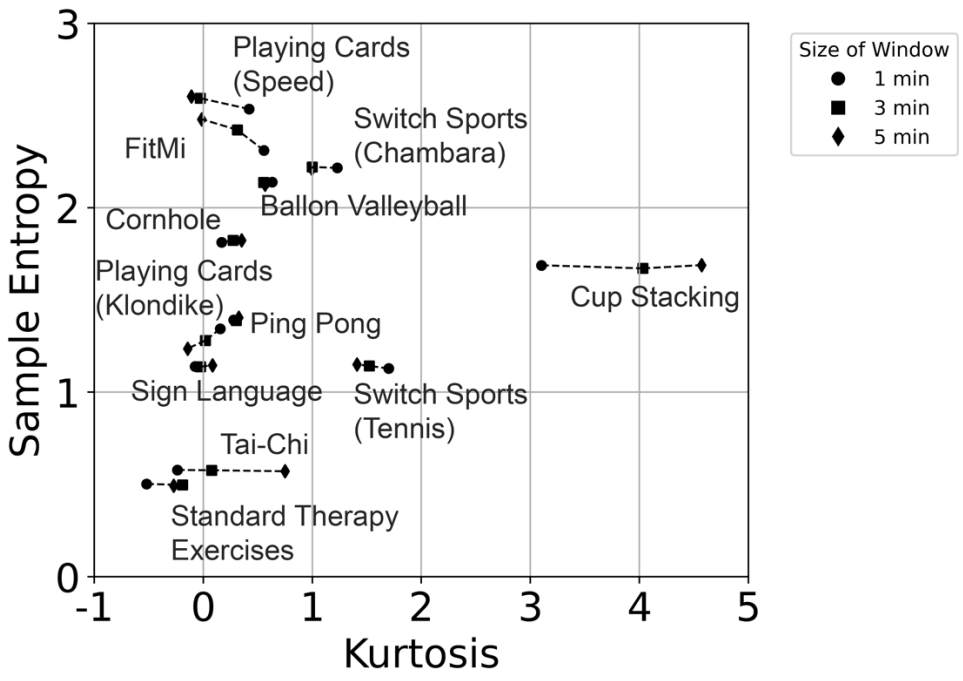



*Figure 17. The mean of kurtosis and sample entropy varying the window length from 1, 3, 5 minutes from 12 activities.*

In order to analyze the between-participant variability in complexity and diversity for each activity, we used box plots for both sample entropy and kurtosis (Figure 18). For sample entropy, certain tasks, such as playing cards speed, switch sports chambara, and FitMi, displayed a larger interquartile range, indicating higher, between-participant variability in the data. In contrast, other tasks such as Tai-Chi and standard therapy revealed more consistent values with smaller interquartile ranges, suggesting a more uniform behavior across participants. For kurtosis, switch sports (Tennis and Chambara) and cup stacking, displayed a larger interquartile range, indicating higher variability in the data. In contrast, other tasks such as standard therapy revealed more consistent values with smaller interquartile ranges.

With the increase in the time window size from 1 to 5 minutes, we observed small changes in interquartile ranges for most tasks. However, some tasks, such as Tai-Chi and Cup Stacking, demonstrated a noticeable change depending on the window size. This finding implies that the complexity and structure of the underlying signals may mildly influenced by the task's duration across the window lengths we studied.

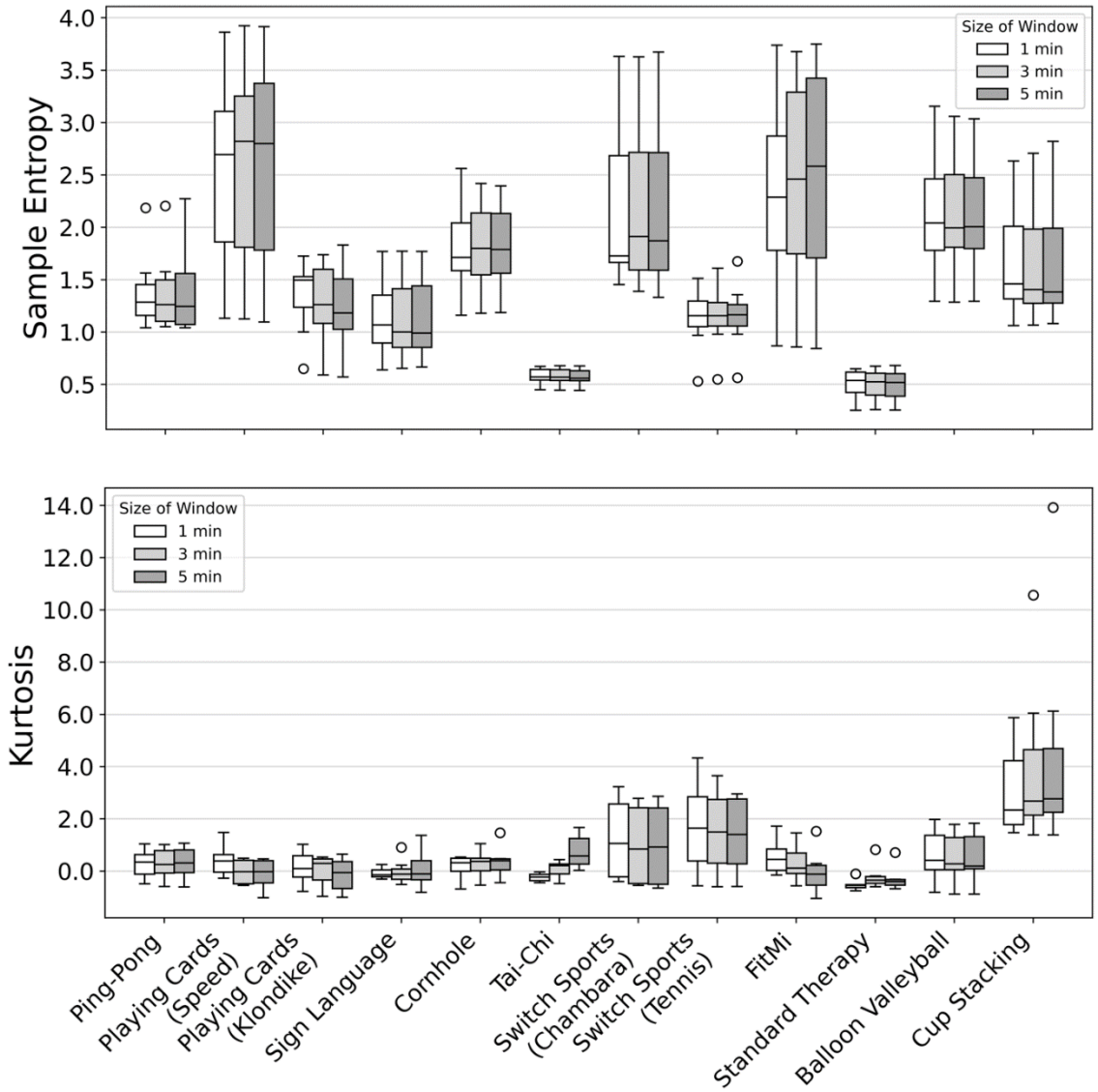


*Figure 18. Boxplots illustrating the distribution of sample entropy and kurtosis of tilt angle for 12 different tasks at 1, 3, and 5-minute time windows across participants (i.e., one point per participant). The box color represents the window size: white (1 min), light gray (3 min), and dark gray (5 min). The solid black line within each box represents the median value, while the whiskers extend to the minimum and maximum values within 1.5 times the interquartile range.*

### *4.4. Discussion*

This experiment aimed to investigate the extent to which various activities promote forearm postural diversity and complexity in unimpaired individuals. The participants engaged in each activity for 10 minutes, wearing a wrist IMU on their dominant hand. The study analyzed the sample entropy and kurtosis of the tilt angle using sliding windows of 1, 3, and 5 minutes. The results indicated that different candidate activities caused movement experiences with differing levels of movement diversity and complexity. Further, diversity and complexity were consistently estimated with all three window sizes. In the following, we discuss these results as well as limitations and future directions.

#### *4.4.1. Interpretation of movement complexity and diversity depending on exercises*

The conventional rehabilitation therapy exercises we selected generated high values for forearm postural diversity but not for complexity. In contrast, playing the card game Speed and conducting exercises with the FitMi exercise system produced the highest values for both postural diversity and complexity. Both Speed and FitMi require the player to try to move as fast as possible. The broader range of speeds experience with these activities compared to conventional rehabilitation therapy may have resulted in

higher complexity values. Consistent with this concept, seated Tai-Chi, which is performed slowly and deliberately, also had low complexity values.

Not all sensor-based exercise were equal however: playing sports with the Nintendo Switch created relatively low values of postural complexity and diversity although playing the game Chambara did exhibit a high complexity. Low complexity with the Nintendo Switch sports may be due to fact that the Nintendo switch likely limits the speed and range of movement required with the joystick controller in order to prevent repetitive stress injury.

Cup stacking showed the lowest diversity of any activity as measured by the high kurtosis values. This makes sense because it required use of a repetitive reach/grasp motion without much variability.

### *4.4.2. Consistency of metric estimates across participants and window lengths*

An important finding was that different activities produced relatively consistent values of diversity and complexity across different participants. This supports the idea that diversity and complexity are inherent to activities, rather than solely dependent on the way activities are performed. Thus, in a wearable feedback scenario, prescription of different activities does seem like a viable approach to challenge the patient to increase diversity and complexity. That said, different activities did have different levels of variability (with Playing Cards and doing FitMi exercises having the highest inter-subject variability), indicating a contribution of the way individuals performed the activities to these metrics as well.

The computed values of diversity and complexity were consistent across different window lengths (1, 3, and 5 minutes) as well, where the window length determined the amount of data that went into the kurtosis and sample entropy calculations. The fact that consistent metrics of diversity and kurtosis can be obtained with a 1 minute window is promising in the context of real-time feedback, where it would be desirable to alert the person when their activities are becoming diverse and complex.

### *4.4.3. Limitations*

This study has several limitations. First, our sample size (n=7) was limited and consisted of only unimpaired individuals; we should confirm our findings by obtaining data from people with a stroke. Second, we only evaluated a limited set of therapeutic activities (n=12). Additional research on a larger set of activities is encouraged to determine the most effective therapeutic activities for promoting forearm postural diversity and complexity in stroke patients. Our findings underscore the potential value of integrating movement complexity and diversity into stroke rehabilitation, especially for UE impairments for people after stroke. We suggest that, moving forward, a formal examination of movement complexity and diversity in guided occupational therapy could yield significant insights. It could be interesting to investigate the differential impacts of low versus high complexity exercises on patient outcomes. In line with this, further data analysis should include analyzing age and handedness dependency.

Further research is needed to determine the optimal design of feedback systems for the quality of the movement experience (QOME). We speculate that users would favor a shorter length of window for analyzing and receiving feedback on the QOME, as it could

let users perceive the change quickly and assign credit to the activity they are doing. A long analysis window would make it difficult for a user to perceive the change of QOME due to their specific activities. The present results suggest that a 1 minute feedback is feasible, and future work should test if even shorter windows are possible. Most importantly, future work should test the hypothesis that quantitative assessments of forearm posture diversity and complexity are beneficial in designing wearable feedback systems for post-stroke rehabilitation, because they can encourage patients to practice specific exercises or activities that challenge the patient in a desirable way.

# Chapter 5. Tools for implementing and validating algorithms for movement complexity and diversity: quality time and a robotic simulator

## *5.1. Summary of the Chapter*

This chapter presents work toward implementing a real-time diversity and complexity detection system to provide wearable feedback to people with stroke about the quality of movement experience (QOME) of the upper extremity (UE). First, we propose the concept of Quality Time (QT) to provide feedback about QOME. QT is a measurement of the number of seconds each day that UE diversity and complexity exceed threshold values. We propose a method to choose appropriate thresholds for each wearer, based on a clinical measurement of their UE impairment level (the Upper Extremity Fugl-Meyer score) and the Minimal Detectable Change (MDC) values for kurtosis and sample entropy measured across a sample of 22 people with stroke. Second, we describe the design and initial testing of a robotic simulator for validating QOME feedback algorithms. The robotic simulator incorporates dual servo motors to emulate human arm supination/pronation and elbow flexion/extension. Experiments were conducted using the test bench to implement movement sequences with known values of complexity and diversity. We then quantified how well an IMU worn on the robot wrist measured these values of complexity and diversity.

*5.2. Introduction*

Real-time feedback is a valuable tool in physical rehabilitation after stroke, providing patients with immediate information on their performance, allowing them to make necessary adjustments and improve their overall rehabilitation outcomes. In the previous two chapters, we identified potential metrics for providing real-time feedback on the quality of the movement experience (QOME) of the upper extremity after stroke. These chapters showed that diversity (measured with kurtosis) and complexity (measured with sample entropy) of the forearm tilt angle can be used for assessing QOME during activities that might be included as part of clinical or home rehabilitation. The next step to enable real-time QOME feedback is to embed the diversity and complexity metrics on a wearable sensor, since, to this point, all metrics were calculated off-line. This chapter discusses work toward implementing and validating the diversity and complexity algorithms on a wrist-worn IMU.

First, we propose the concept of Quality Time (QT) to provide real-time feedback about QOME. We design a patient-adaptive threshold strategy that sets a threshold for counting time spent in movements with high-than-normal QOME, based on the user's UEFM score and the Minimal Detectable Change (MDC) values for diversity and complexity. Second, we present the design and initial validation of a robotic simulator capable of emulating human forearm supination-pronation and elbow flexion-extension. This simulator allows us to implement movement patterns with known complexity and diversity, and then assess the accuracy of a wrist-worn IMU (mounted on the robot wrist) and associated algorithms to estimate these values.

*5.3. Methods*

*5.3.1. An Adaptive Goal Setting Strategy for Measuring Quality Time*

We first focused on developing a way to calculate the duration of time that users of the wearable sensor are engaging in high quality movement experiences, a variable we call “Quality Time” (QT). We were inspired by Demers et al., who previously devised an Active Time (AT) algorithm that computes the amount of time the UE is active, without regard for movement quality. They demonstrated a positive correlation between AT and the number of arm movements that participants with a stroke performed throughout the day [37]. We propose QT as a feedback metric because it is easily interpretable by patient-users, similar to AT.

The QT algorithm we propose here quantifies the amount of time when UE movement is deemed sufficiently complex and diverse using user-specific thresholds defined in such a way as to identify periods of time when complexity and diversity are “greater than normal”. The reason it is necessary to assign thresholds on a user-specific basis is that different users have different levels of movement ability. If a threshold for assessing periods of QT is too high, then users who are more severely impaired will not be able to make movements that exceed this threshold and increase their amount of QT. Conversely, if the threshold for assessing QT is too low, then all movement throughout the day will contribute to QT, insufficiently challenging the user.

To establish optimal rehabilitation goals, we propose a user-adaptive threshold-setting strategy based on clinical scores and historical IMU data obtained from people with a stroke who wore the IMU during daily activities. This strategy allows us to create

goals that are adapted to each individual's impairment level, aiming to keep patients engaged and motivated throughout their at-home rehabilitation.

We set user-specific goals for sample entropy and kurtosis using the same dataset that we used in Chapter 3, obtained from a pilot clinical trial of wearable feedback which contains the movements of 22 participants during their daily activities recorded with a wrist-worn IMU (the Manumeter) [70]. First, we computed the sample entropy and kurtosis for each of three groups of participants within the dataset, where the groups were defined by the level of UE impairment (severe, moderate, mild, see Chapter 3). We defined the MDC using a confidence interval formula:

$$CI = \bar{x} \pm z\frac{s}{\sqrt{n}} \tag{12}$$

where x = sample mean, z = confidence interval, s = sample standard deviation, and n = sample size. We calculated MDCs for the three groups (n = 7, 12, and 15, for severe, moderate, and mild groups, respectively) and then fit a second-order polynomial curve to the data in the range of the UEFM score from 15 – 60. We adjusted z from MDC50 to MDC99, dividing it into 100 elements, where MDC50 and MDC99 represent the MDC with a statistical power of 0.5 and 0.99, respectively, to detect the change.

### *5.3.2. The Robotic Simulator for Algorithm Validation*

A critical aspect of implementing QOME algorithms is the absence of ground truth. We can perform calculations to estimate kurtosis and complexity from IMU data obtained from human subjects, but how do we know these calculations are accurate? Validation becomes especially important when implementing the calculations in real-time embedded software. To address this validation challenge, we developed a robotic

simulator that allows us to create human-like movements with known levels of movement complexity and diversity.

The robotic simulator incorporates two servo motors (FS5103B, FEETECH, China) configured to simulate human forearm supination/pronation and elbow flexion/extension (Figure 19). The simulator is composed of the following components: 1) a dual servo motor system, which uses two servo motors to generate rotational motions during experiments, achieving the desired movement complexity; 2) an ABS arm, which connects the motors or actuators to the platform in a configuration that mimics the human forearm; 3) a wrist-worn IMU can be attached to the arm, just like to a human forearm; 4) a 3D printed ABS platform which supports the servo motors enabling rotational movements; 5) ABS frames are fastened to a wooden board with screws for added stability; 6) a microcontroller (Arduino Uno, Italy), which controls the dual servo motor system during tests. The arm and the platform were both printed by the 3D printer (Flashforge Pro, China). The simulator was clamped onto a bench-top table to ensure stability during experiments.

### *5.3.3. The Wrist-Worn IMU*

To estimate the posture of the robotic forearm and calculate complexity and diversity, we used the MiGo, a wrist-worn inertial sensor developed by Flint Rehabilitation Devices LLC (Irvine, California, USA). The MiGo has a six-degrees-of-freedom (DOF) inertial measurement unit (IMU, LSM6DSL, accelerometer range set to ± 2 G and gyroscope range set to ± 500 degrees per second, both with a 16-bit resolution). Time and date are managed by a real-time clock (PFC2123). A system-on-

a-chip (NRF52, Nordic Semiconductor) with an ARM Cortex M4 CPU and an integrated 2.4gHz radio is used to read and handle the data from the IMU. The MiGo supports a streaming mode, in which sensor data can be streamed at 100Hz using the enhanced shock-burst wireless protocol. Alternatively, it can be operated in tracker mode, in which it runs an algorithm to track upper extremity active time.

### *5.3.4. Assessing Movement Complexity and Diversity*

We utilized sample entropy and kurtosis algorithms to evaluate and analyze the movement quality of the robotic simulator. We employed tilt angle as a variable, calculating the angle between the normalized acceleration vector in the sensor coordinate frame and the gravity vector (see Chapter 3), inspired by Leuenberger et al [121]. Sample entropy, a measure quantifying the regularity and complexity of time series signals, was computed using a formula that involved the number of matches of length m and m+1 with their respective templates (we chose m to be 3; see Chapter 3). Kurtosis, a statistical measure representing the "tailedness" and "sharpness" of a probability distribution, was determined by dividing the fourth central moment ($\mu_4$) by the standard deviation (σ) raised to the fourth power. The fourth central moment and standard deviation were calculated using their respective formulas. By computing kurtosis and sample entropy, we repeatedly evaluated the movement diversity and complexity of the robotic simulator as it moved. We chose the evaluation window to be 1 minute in duration based on the results in Chapter 4, sliding the window by 1 second intervals to produce continuous measures of diversity and complexity based on the previous 1 minute of IMU data.

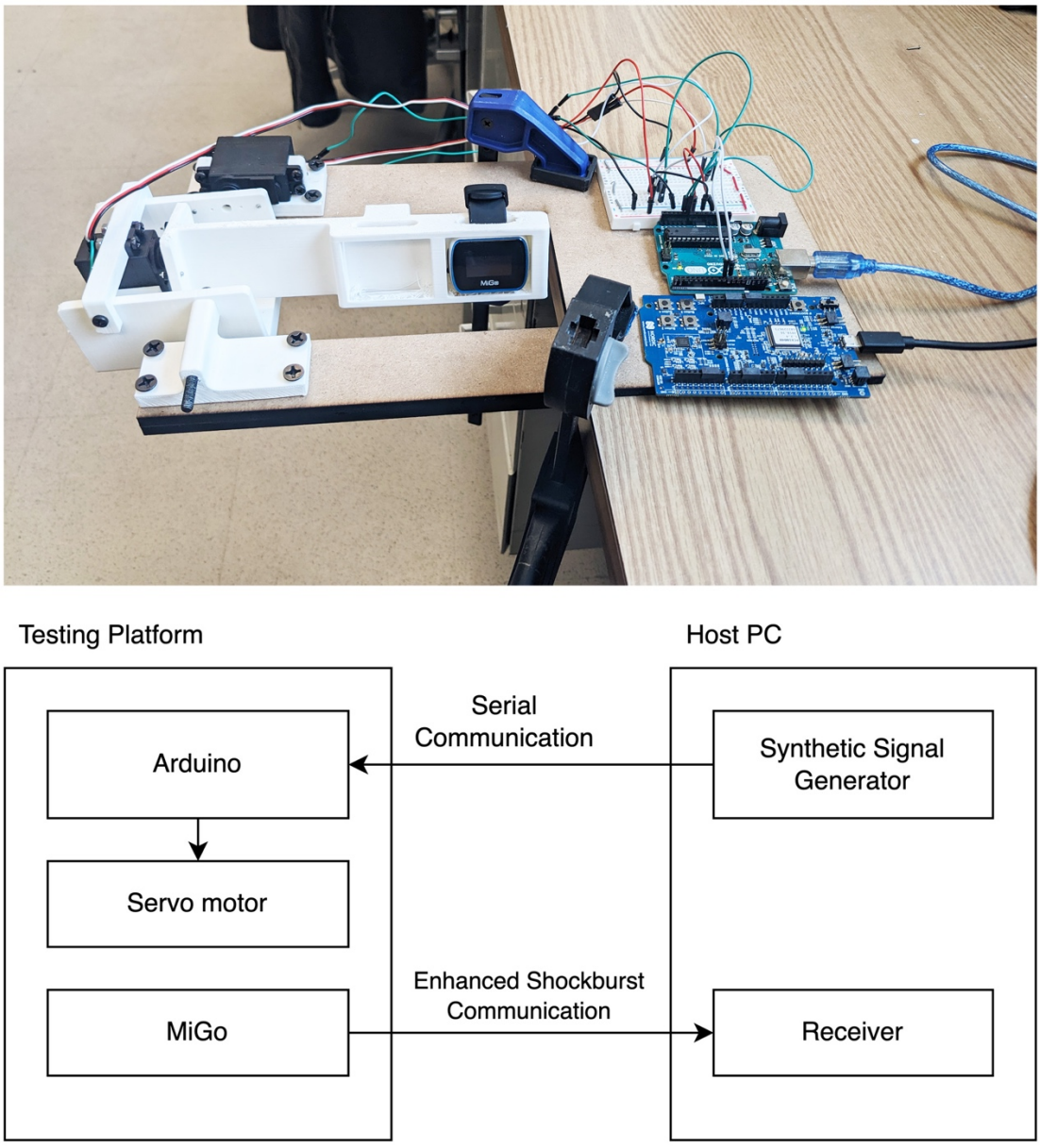



*Figure 19. Top: The robotic forearm simulator as a test bench to validate wearable sensing algorithms for diversity and complexity. The simulator emulates elbow flexion/extension and forearm supination/pronation. Bottom: Schematic of control system for the robotic simulator. For validating kurtosis, the distribution generator was used. For validating sample entropy, the synthetic signal generator was used.*

*5.3.5. Experimental protocol*

Experimental protocols using the robotic simulator were designed to validate the ability of the wrist-worn IMU to estimate kurtosis and sample entropy. The servomotor simulating arm supination/pronation was programmed to move along trajectories with various diversity and complexity using the microcontroller. The robot pronation/supination angle varied from 0 to 180 degrees to match the range of the human forearm.

The following hierarchical control architecture was used. The microcontroller waits in standby until it receives a command from the host PC. Once it starts, the host PC uses a Python program to send a movement instruction to the microcontroller, which in turn drives the servo motor to execute the specified movement. After the microcontroller receives the instruction, it sends a confirmation message back to the Python program. If the Arduino does not send a response within a specified timeout, the computer will attempt to resend the combination up to a predetermined number of times. This architecture allows the movement sequences to be easily modified on the host PC by editing the Python program.

For validation of the kurtosis measurement, three sets of joint angle targets with different kurtosis values were generated: uniform, normal, and Laplace distributions (Figure 20). Each set contained 250 joint angle targets. The robot was programmed to execute each targeted movement within a span of 3-4 seconds each. This resulted in a total experimental duration of approximately 10 minutes, during which the robot cycled through the targets, returning to the initial resting position between each movement.

The movements were orchestrated using a minimum jerk trajectory model, which mimics smooth, human-like motion trajectories [154]:

$$x(t) = 6t^5 - 15t^4 + 10t^3 \quad (13)$$

where t is time, and x is the function of position. Kurtosis was calculated for each dataset and compared to the known kurtosis value of the distribution used to generate the joint angle target set.

For validation of the sample entropy measurement, target joint angle trajectories were generated to vary sample entropy values, and the servo motor was programmed to try to follow these trajectories. Three trajectories based on three types of noise with different entropy levels were created: brown noise, pink noise, and white noise (Figure 21).

During the robot movements, the MiGo streamed IMU data to a desktop computer at 100 Hz.. We down-sampled signals offline from 100 Hz to 52.6 Hz using the scipy.resample_poly function, in order to match the sampling rate we used in Chapters 3 and 4.

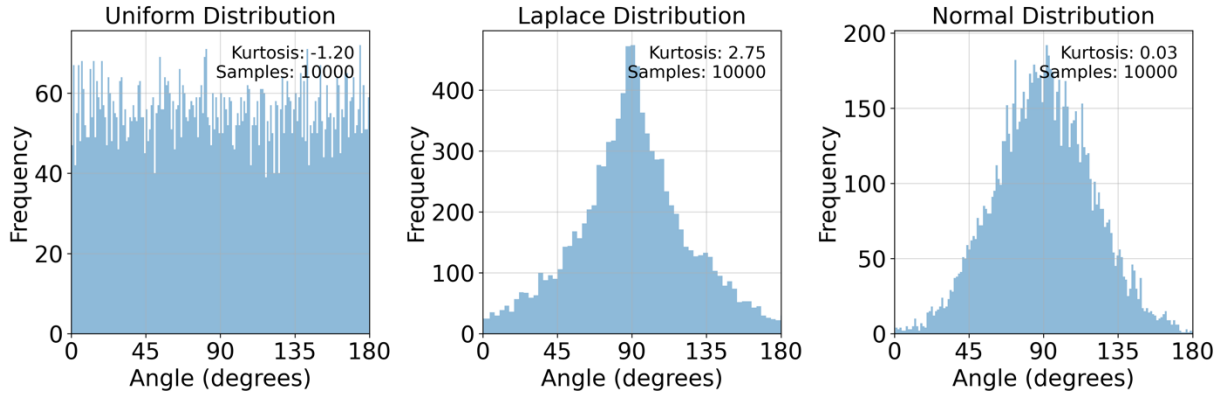



*Figure 20. Examples of three target distributions of robotic forearm posture (Uniform, Laplace, and Normal) for validation of the diversity measure (i.e. Kurtosis).*

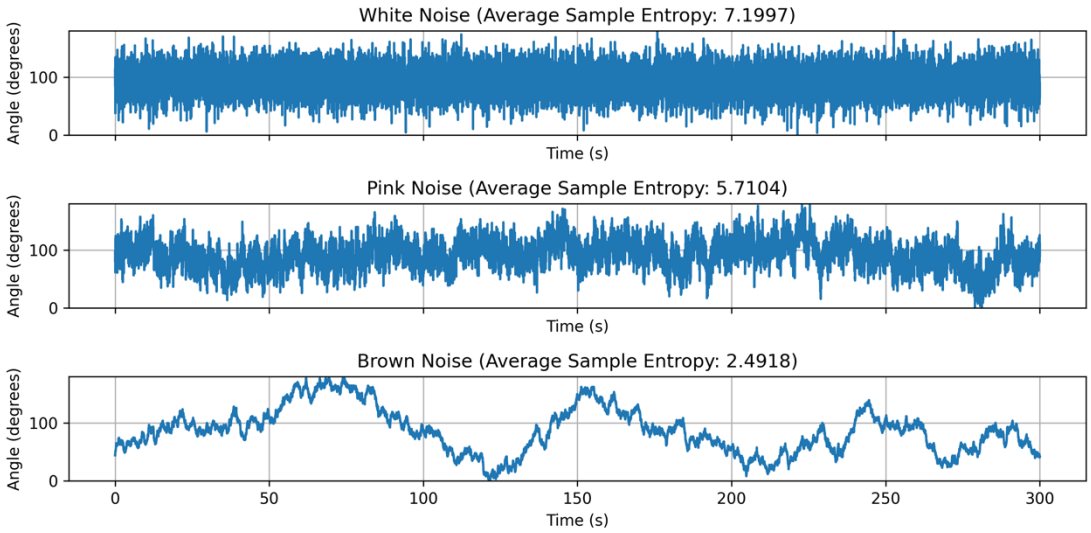



*Figure 21. Example of target forearm postural trajectories (White, Pink, and Brown Noise) used for the validation of the complexity metric (i.e. SampEn).*

### *5.3.6. Data Analysis*

In some cases, we applied a Butterworth low-pass filter on both the simulated data and actual data, using a 5 Hz cutoff frequency. We calculated the tilt angle of the robotic forearm as the arccosine of the normalized filtered acceleration vector (see Chapter 3). We calculated the distribution of the tilt angle signals for both the planned and actual trajectories. To evaluate differences between planned and experimental trajectories, we calculated the median of the sample entropy and kurtosis of the tilt angle signals. To further assess the discrepancies between the planned and experimental, we applied the Manhattan distance. The Manhattan distance $d_M$ is calculated as follows:

$$d_{\mathrm{T}}(\mathbf{p}, \mathbf{q}) = |\mathbf{p} - \mathbf{q}| = \sum i = 1^{n} |pi - qi| \quad (14)$$

In $\mathbb{R}^2$, $d_M$, between two vectors $\mathbf{p} = (p_1, p_2, \ldots, p_n)$ and $\mathbf{q} = (q_1, q_2, \ldots, q_n)$ in an n-dimensional real vector space is the sum of the lengths of the projections of the line segment between the points onto the coordinate. Here we employed $d_M$ having two dimensional vectors $\mathbf{p} = (p_1, p_2)$ and $\mathbf{q} = (q_1, q_2)$. The difference was computed between the point where $(p_1, p_2)$ = (the median of the simulation data, and the median of the actual data), and the closest point $(q_1, q_2)$ on the y = x. This point on the y = x line represents the ideal condition if there is no bias between the simulation data and the actual data.

We also utilized Power Spectral Density (PSD) analysis to examine the frequency content of the tilt angle signals obtained from the robotic simulator, using Welch's method. The PSD is a widely used method in signal processing and analysis for characterizing the frequency content of a time series signal. It represents the distribution of power over various frequency components, providing valuable insights into the dominant frequencies present in the signal and the relative contribution of each frequency to the overall signal. We compared the PSD of the actual data and the simulation data, allowing us to gain insights into the dominant frequencies and any noise present in the datasets. Subsequently, we computed the Mean Squared Error (MSE) to measure the difference between the PSDs of the actual and simulation data.

### *5.4. Results*

#### *5.4.1. Estimating Quality Time*

We used wrist-worn IMU data recorded from 22 people with a stroke during daily life to estimate user-specific thresholds for identifying periods of movement with high-

than-usual complexity and diversity. The thresholds for sample entropy and kurtosis are shown in Figure 22, plotted as a function of UE impairment level and the statistical power required to designate a change as significant. As UE impairment becomes less severe, the thresholds become more challenging to exceed.

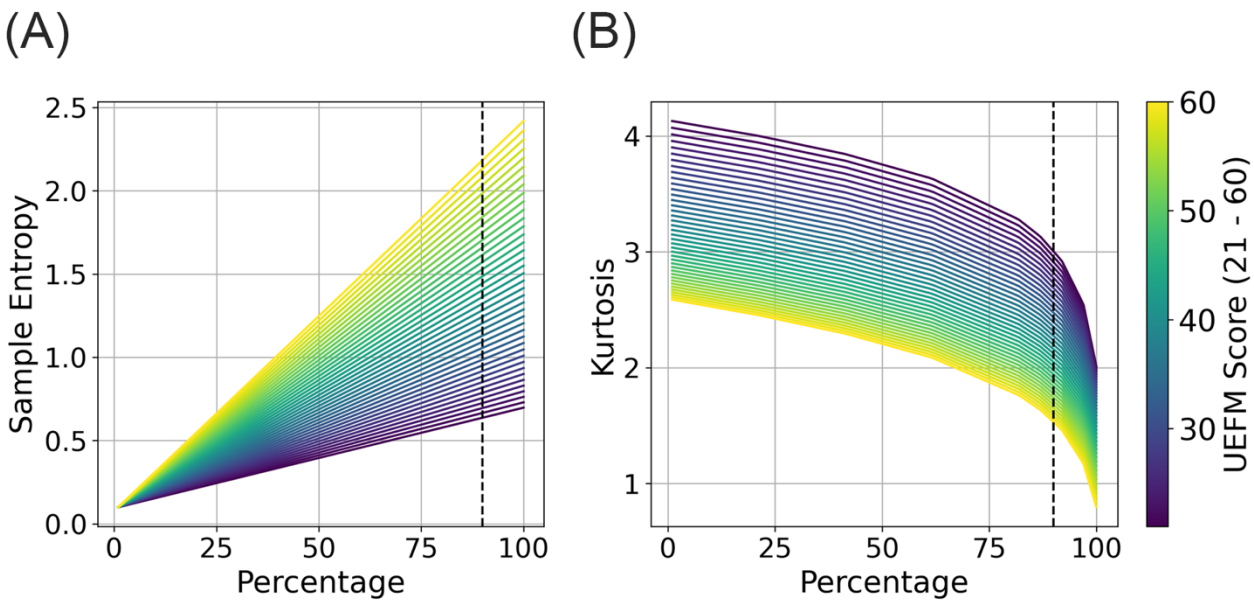



*Figure 22. The proposed, user-specific, goal setting strategy for the sample entropy and kurtosis thresholds. The idea is to choose threshold values for sample entropy and kurtosis for each user, based on their UEFM score. The colored lines show the Minimal Detectable Change for sample entropy (left) and kurtosis (right) at different levels of statistical power (denoted by "Percentage" on the x-axis), and for different levels of UE impairment (denoted by the line color). We set the user-specific threshold based on the MDC90, denoted by the vertical dashed line (i.e the MDC with 0.9 statistical power). When the sample entropy and kurtosis, calculated over a time window, exceed this threshold, then the Quality Time counter is incremented. For sample entropy, more severely impaired users have a lower threshold, since a higher sample entropy denotes more complex movement. For kurtosis, more severely impaired users have a higher threshold, since lower kurtosis values denote a more diverse movement experience.*

We also used the historic wearable sensor data set obtained from the 22 participants with a stroke to estimate how many seconds of QT were achieved over a 12 hour period for people with different impairment levels, using this user-adaptive threshold strategy. Each user achieved a daily QT between 500 and 10000 seconds (Figure 23). Thus, defining a user-adaptive threshold made it so no user had 0 seconds of QT, and no user had 43,200 seconds of QT, which would occur if all 12 hours of wear time were counted.

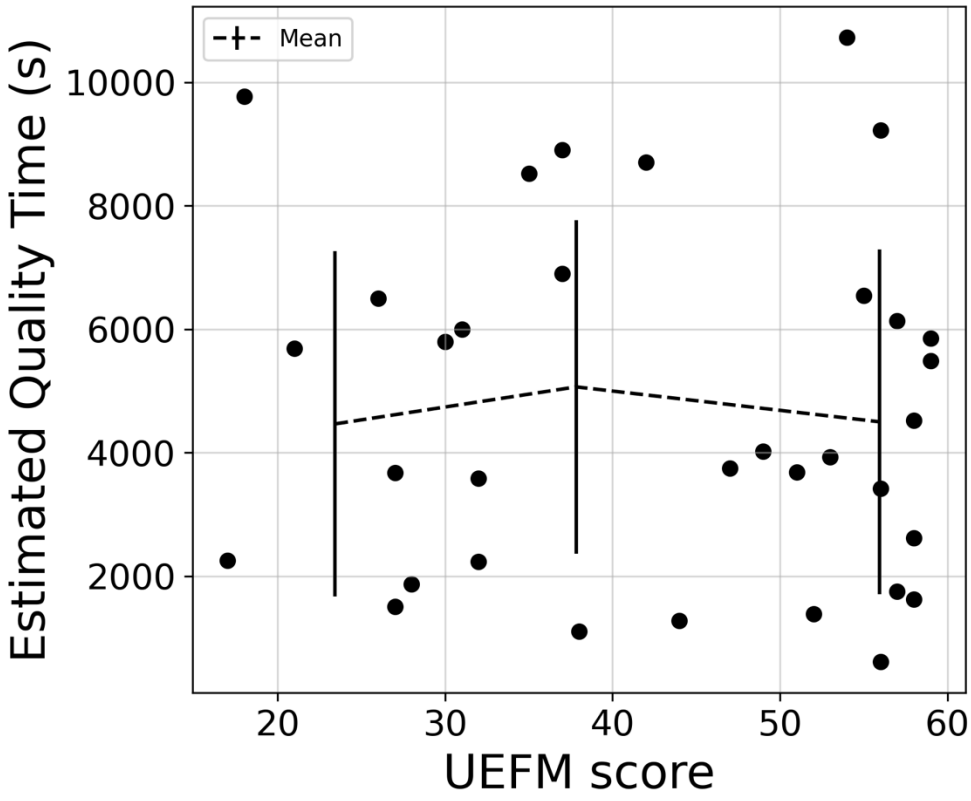



*Figure 23. Estimated Quality Time (QT) over a 12 hour period as a function of UEFM score for 34 data sets of IMU recordings obtained from 22 people with stroke who wore the IMU on their wrist in daily life. The QT calculation uses the user-adaptive threshold.*

*5.4.2. Measuring complexity and diversity of movements generated with a robotic forearm*

We observed differences between the planned robotic forearm angle distribution and the experimentally measured distribution, particularly for the pink and white noise distributions (Figure 24). The possible sources of error were: 1) the robot did not have high enough bandwidth to accurately track the planned tilt angle distributions; and 2) the estimation of tilt angle from the IMU data was inaccurate.

To further investigate the bandwidth issue, we analyzed the power spectral density (PSD) of the tilt angle (Figure 25). Before applying the 5 Hz low pass filter, we observed a large mean squared error and a more considerable distance between the distributions. This indicated that the robot did not have sufficient bandwidth to keep up with the higher frequency components in the planned signals. Further analysis therefore focused on signals after they had been filtered with the 5 Hz low pass filter.

Upon comparing the tilt angle distributions between the planned and experimental trajectories after low pass filtering, we also found that the differences in medians were all larger than 5.0 degrees. This suggest there was an offset between the forearm angle at rest in the simulation and the experiments (Figure 24).

The experimentally measured kurtosis and sample entropy values, based on the signals from the IMU mounted on the robotic forearm, with the planned kurtosis and sample entropy values were compared (Figure 26). Recall, these values were calculated using a 1 minute window sliding in one 1 second increments across 10 minutes of robotic movement. As the planned kurtosis and entropy increased, the experimentally measured kurtosis and entropy increased, although there were some

inaccuracies, particularly for the experimentally measured entropy for the pink and white noise. This is likely due to the inability of the robot to keep up with the high frequency content of the pink and white noise.

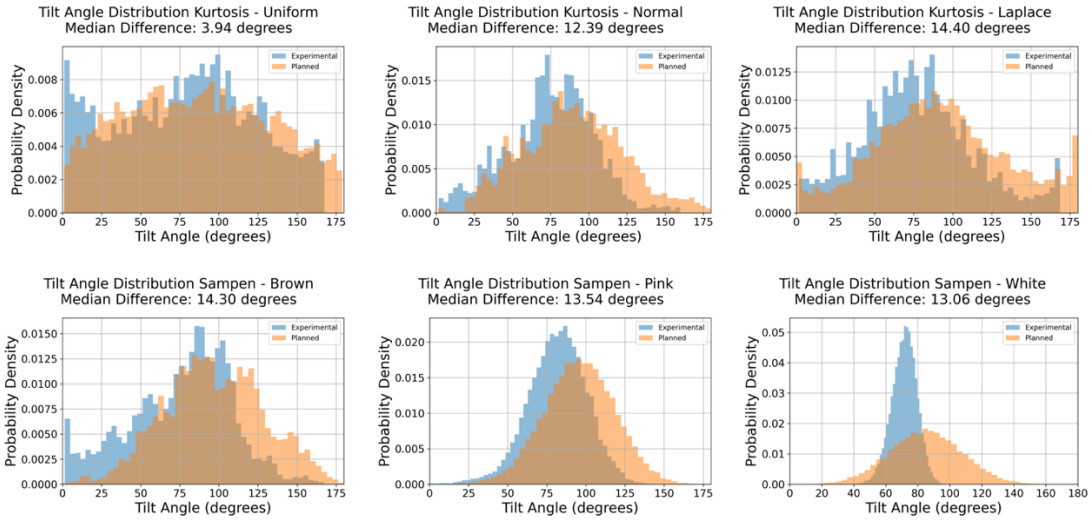



*Figure 24. Comparison of planned and experimental tilt angle after low-pass filtering. The differences in the median values were all larger than 5.0 degrees, suggesting an offset between the angle at rest in the simulation and experiments.*

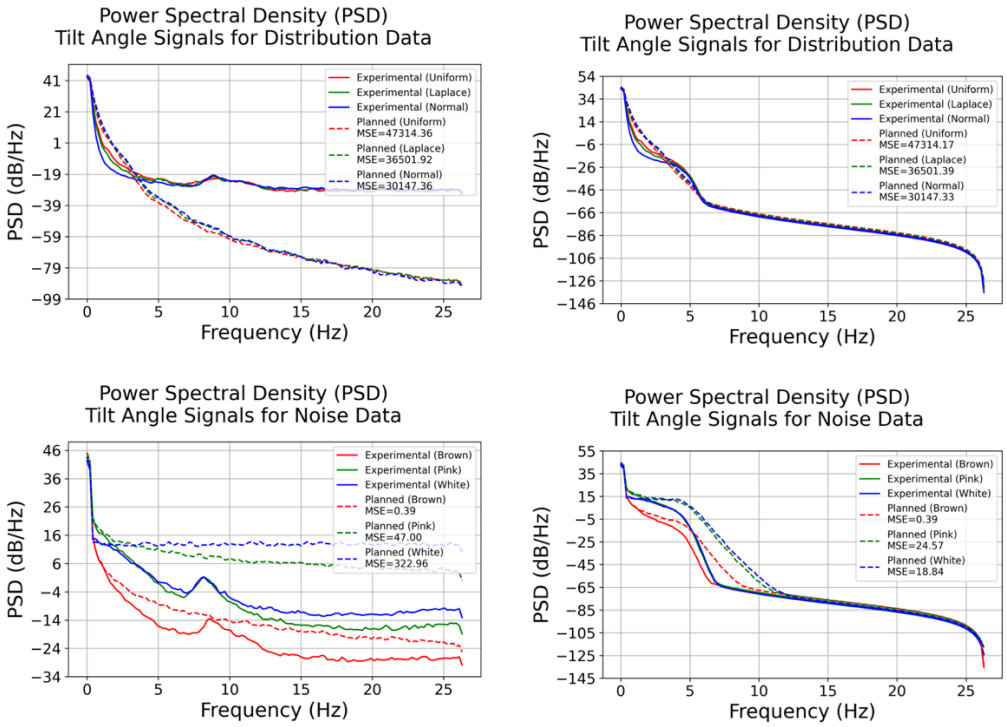



*Figure 25. Comparison of Power spectral density (PSD) of tilt angle signals for the planned and experimental trajectories. (Left) Before 5 Hz low pass filter was applied, (Right) After 5 Hz low pass filter was applied*

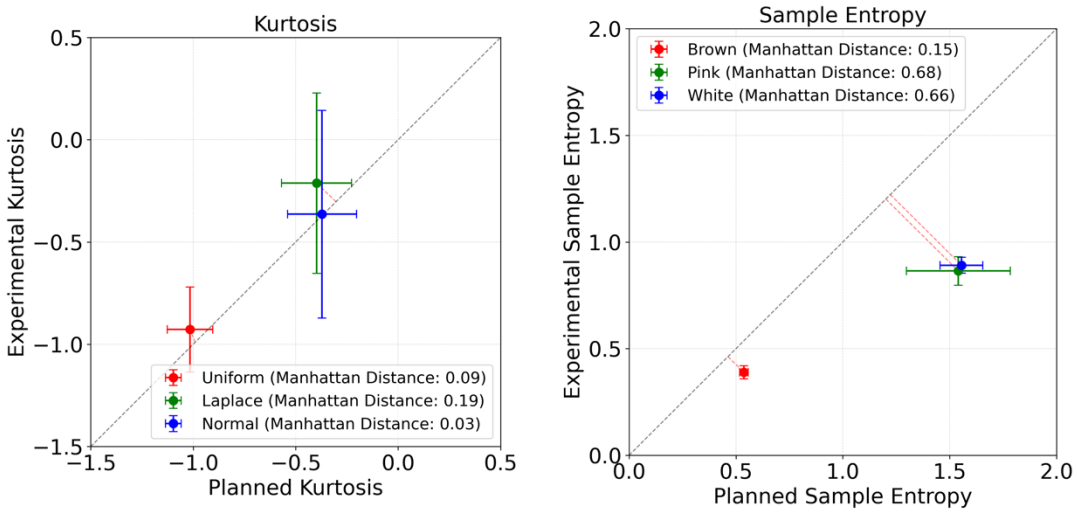



*Figure 26. Left: Mean values for the experimentally measured entropy compared to the simulated (or "theoretical") entropy. Error bars show ± one standard deviation for entropy and kurtosis evaluated over 1 minute windows. Right: Experimental and simulated kurtosis.*

### *5.5. Discussion*

In this chapter, we proposed and evaluated two tools for assisting in real-time implementation of Quality of Movement Experience (QOME) feedback. First, we proposed Quality Time as a feedback metric for quantifying time when subjects are performing higher-quality-than-usual movements. The adaptive threshold for movement quality scoring was based on the UEFM score and MDC values derived from distributions of sample entropy and kurtosis obtained from 22 persons with a stroke in previous study. We also designed a robotic simulator as a tool to validate the metrics of QOME obtained from a wearable sensor. The simulator allowed us to create movements similar to human forearm movements, but with planned and varying values of kurtosis and complexity. This allowed us to validate the measures of kurtosis and complexity calculated from experimental data obtained with the wrist-worn IMU,

although there were some inaccuracies still to be overcome. We briefly discuss these results, including limitations and directions for future work.

### *5.5.1. Quality Time as a goal setting strategy*

We introduced a user-adaptive goal-setting strategy based on patients' functional capacities and utilizing MDCs of kurtosis and entropy estimated from a data set of 22 people who wore an wrist-IMU in daily life. The idea of Quality Time aligns with the non-task-specific training philosophy proposed by Krakauer et al. [66], emphasizing the exploration of more complex and diverse movements for stroke rehabilitation. To refine this goal-setting strategy, future research could explore the impact of sample size on the confidence interval calculations.

### *5.5.2. A robotic simulator for validating embedded quality of movement measurements*

The robotic simulator enabled us to validate the proposed algorithms for measuring QOME by providing a controlled environment for testing arm movements. Specifically, it allows us to pre-program planned trajectories with known values of complexity and diversity, and then check that our algorithms can reproduce these values based on IMU recordings. However, we experienced difficulties with having the robot reproduce the planned trajectories, due to limited bandwidth of the motor and a bias in the motor angle. These issues will be addressed in future experiments by ensuring that the planned trajectories are achievable by the robotic simulator, and by better calibration of the robot simulator angles.

### *5.5.3. Future work*

Future work should implement and assess the proposed goal-setting strategy using the robotic forearm simulator. The effectiveness of the quality time concept can be validated using the robotic simulator to design the optimal difficulty level for simulated users. By tailoring the difficulty and goal setting to individual users, we can contribute to improved stroke rehabilitation outcomes and personalized therapy plans.

# Chapter 6. Conclusion

The aim of this dissertation was to develop new algorithms that can improve the effectiveness of wearable technology for UE rehabilitation for people after stroke. This chapter summarizes the contributions and discusses future work for the use of wearable sensors for people after stroke.

In Chapter 1, we reviewed key concepts in stroke rehabilitation and therapies using wearable sensors. We discussed existing rehabilitation technologies and emphasized the problem of non-use in stroke patients, who often fall into a vicious cycle [18] of not using their impaired arm, leading to a decline in UE function. We also discussed the potential of wearable sensors to provide feedback and promote the therapeutic use of the impaired arm in daily life.

In Chapter 2, we described the development of the HARCS algorithm, a convolutional neural network trained on spectrogram data from a wrist-worn IMU to identify hand movement occurrence. We found that counts obtained using the HARCS and the previously-developed HAND algorithm that requires magnetic sensing were highly correlated. This indicates that is possible to identify hand movements of people after stroke in daily activities without an external source of information such as a magnetic ring on the index finger. However, the HARCS is not perfect, suffering from both false positives and negatives.  It is unclear if this is a fundamental limit of the approach or whether HARCS could be further improved by refining the strategy. Regardless, and important contribution of this dissertation is to show for the first-time how wrist-worn IMU data can be used to estimate the amount of finger movements,

opening up new avenues of existing wearable watches for hand-related healthcare applications.

In Chapter 3, we investigated what statistical variables related to the quality of movement experience (QOME) distinguished persons with a stroke impairment level to understand differences in movement quality between mildly, moderately, and severely impaired patients. We identified kurtosis and sample entropy (SampEn) of tilt angle are having a highest effect size (Cohen's D = 1.1, and 0.99 for group 1 vs group 2, and group 2 vs group 3, respectively), suggesting they can be used as a measure of movement diversity and complexity. Furthermore, we utilized a variable selection method [133] with a backward elimination to identify the optimal set of variables for predicting UEFM score. We again obtained kurtosis and sample entropy as suggested variables. Identifying metrics of the quality of the movement experience is a novel contribution of this dissertation.

Based on the findings of Chapter 3, we hypothesized that encouraging people to achieve more forearm postural diversity and complexity may be therapeutically beneficial, as it challenges the patient to make more challenging and variable movements. But the following question remained unclear: what exercises could one suggest to help patients achieve this goal? To answer this question, in Chapter 4, we recruited 8 unimpaired individuals and evaluated a set of 12 therapeutic activities for postural diversity and complexity. The activities were performed while seated and included: a set of conventional rehabilitation therapy exercises for the hand and arm, ping-pong, the card game Speed, American Sign Language, Tai-chi, Cornhole, Nintendo Switch sports, balloon volleyball, cup stacking, and FitMi exercises, where

FitMi is a commercial sensor system that guides upper extremity exercises. The participants performed each exercise for 10 minutes while wearing a wrist accelerometer, and we computed sample entropy and kurtosis of tilt angle with sliding windows of 1, 3, and 5 minutes. Engaging in conventional rehabilitation therapy exercises created high values for forearm postural diversity but not complexity. Playing the card game Speed and exercising with FitMi produced the highest values for both postural diversity and complexity. Measures of diversity and complexity were reasonably consistent for each activity, highlighting that they are a property of the activity rather than the way persons perform the activity. These measures were also consistent across 1, 3, and 5 minute data analysis windows, which may be helpful for providing timely feedback to patients when they start to perform diverse, complex activities.

In Chapter 5, we provided an overview of work in developing an algorithm for providing feedback on QOME in a real-time. We designed the algorithms to quantify "quality time" when users are in active in terms of high movement complexity and diversity. We also designed and performed pilot testing with a robotic simulator of the human forearm for validating movement complexity and diversity.

Future work will focus on further developing and refining the algorithms and technologies discussed in this dissertation, as well as implementing them in a robust, user-friendly way in a wrist-worn IMU. For the HARCS algorithm, it may be possible to improve its accuracy with further optimization of parameters or better selection of training data sets. For QOME wearable feedback, important questions to address are: 1) what is the optimal duration for the data analysis window to give consistent but also

timely, real-time feedback on activities? 2) Should we provide feedback on both kurtosis and complexity, or will providing feedback on one metric suffice? Providing feedback on complexity is more computationally expensive than kurtosis, and it is still an open question which metric is more motivating and relevant. 3) Are there other measures of QOME that are relevant? 4) What is the best strategy for providing feedback about diversity and complexity, in terms of frequency and use of visual and haptic cues? 5) How can we combine wearable sensing with remote assessment that is monitored by a rehabilitation professional? 6) Can we develop computational models to predict recovery using wearable sensor data? And 7) Are there potential synergies between QOME wearable feedback and other rehabilitation strategies in post-stroke patients?

Finally, this work lays the foundation for a forthcoming clinical trial to examine the effects of QOME feedback with a wrist-worn IMU. The study will be conducted in an outpatient rehabilitation unit, and will offer the wrist-worn IMU as an "add-on" to patients' therapy. The target population consists of individuals approximately 1-3 months post-stroke. The goal is to enroll 30 stroke patients over a 1.5-year period, randomized into two groups:

1. Quantity feedback: The wrist-worn IMU informs patients of the daily duration of their arm movement.
2. Quality of Movement Experience (QOME) feedback: The wrist-worn informs patients of the daily duration of their "high-quality" arm movement and vibrates when they achieve a 1-minute period of high-quality movement (i.e., good forearm postural diversity and complexity).

Both groups will receive a suggested set of activities, selected from the activities we tested in Chapter 4, including playing cards, balloon volleyball, learning sign language, seated tai-chi, conventional book of exercises, and cup stacking. Each group will receive different instructions for engaging in these activities, with the quantity group focusing on increasing arm movement activity and the QOME group aiming to find activities that trigger the wrist-worn sensors vibration. Both groups will also have personalized daily goals for the number of seconds spent on arm movement. We will test which wearable feedback strategy most improves participant's ability to move their UE (capacity), as well as their use of their UE in daily life (performance).

## Supplementary material

### *1.1. Supplementary material for Chapter 2*

#### *1.1.1. Inactivity Filtering*

In *the Manumeter-Home Dataset* and *Manumeter-Lab Dataset*, there existed significant periods of inactivity during which all measurements remained approximately constant. Participants most likely took off the Manumeter during this time, or rested with their hand on a table or their lap, suggesting it was inappropriate to assess the network's performance on these intervals. We excluded windows where the inactive regions were detected, where we defined an inactive sample as a time-sample during which the difference between the minimum and maximum values of all 9 measurements - the x, y, z components of the acceleration, angular velocity, and gravity direction - failed to exceed their respective thresholds, thus being approximately constant. The 9 thresholds were calculated based on 4 of the data files in the *Manumeter-Home Dataset*, which were selected due to having long intervals of continuous inactivity. The largest difference in minimums and maximums was computed across the intervals in the 4 files for each of the 9 measurements, yielding the thresholds.

#### *1.1.2. Resampling for Imbalanced Dataset*

Generally, there were significantly more negatively labeled samples than positively labeled samples, which was particularly true for the *Manumeter-Home Dataset* as the participants were not performing specific movements during a pre-set amount of time. For validation sets, we under sampled negatively labeled samples to

weigh each class equally in the calculation of accuracy. We also did this for the training sets, as otherwise the network was biased towards classifying a sample negatively.

### *1.1.3. Transformation to the Frequency Domain*

We computed spectrograms of data-samples by applying a Short-Time Fourier transform (STFT), a signal analysis technique for determining the sinusoidal frequency and the phase content of local parts of a signal. We selected 20 time-samples as the FFT size and utilized a Tukey window with a shape parameter of 0.25. The FFT size for the time sampling rate 52.6Hz produced 2.63 Hz increments in generating a spectrum, resulting in the generation of 11 frequencies from 0 Hz to the Nyquist frequency 26.3 Hz. The windows had an overlap of 19 time-steps in order to convey as much information as possible within the resulting spectrogram. Spectrograms were generated for each measurement type, resulting in each sample containing 9 spectrograms stacked along the last axis, resulting in a sample with a shape of (131 time steps, 11 frequencies, and 9 measurements) (Figure 3).

$$\mathrm{STFT}\{x[n]\}\,(m,\omega) = X(m,\omega) = \sum_{n=-\infty}^{\infty} x[n]\omega[n-m]e^{-j\omega n} \qquad (15)$$

where x[n] represents the signal, and $\omega[n-m]$ represents the adapted Tukey window.

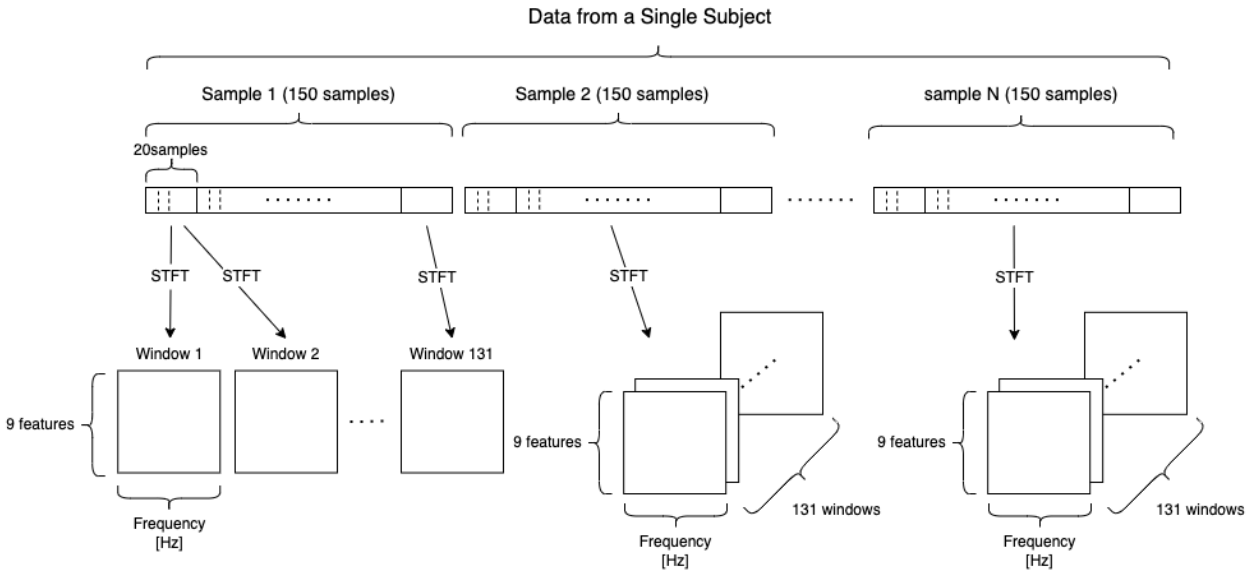



*Figure 27. (Left) An illustration of data processing for a single subject using the Short-Time Fourier Transform (STFT). The STFT creates a (9 features x 11 variables) window to form a spectrogram in 150-time samples. 131 windows were generated from 150 samples.*

### *1.2. Non-Linear Transformation of the Spectrograms*

The generated spectrograms did not have normal distributions in values and had a significantly larger density of low amplitudes. After converting all the samples in the datasets to spectrograms, therefore, we applied a Box-Cox transformation that converted skewed distributions to normal distributions [82], [79]–[81]. Equation 16 shows the formula we used to convert time series signals:

$$x^{\lambda} = \begin{cases} \dfrac{x^{\lambda} - 1}{\lambda} & \text{if } \lambda \neq 0 \\ \log x & \text{if } \lambda = 0 \end{cases} \tag{16}$$

where x and $\lambda$ represent the input value and the parameter varying a scaling of data distribution, respectively. $\lambda$ was chosen independently for every combination of frequency and measurement type, resulting in an array of lambdas with a shape of (11 frequencies, 9 measurements). During the generation of the training set, $\lambda$ was chosen

to maximize the log likelihood function using the SciPy scalar optimizer [76]. Afterwards, the mean and standard deviations (SDs) of the dataset were calculated along the sample and time axes, again resulting in arrays with shapes of (11 frequencies, 9 measurements). The computed mean and SDs normalized the entire dataset. Same $\lambda$, means, and SDs from training sets were used in generating the validation set.

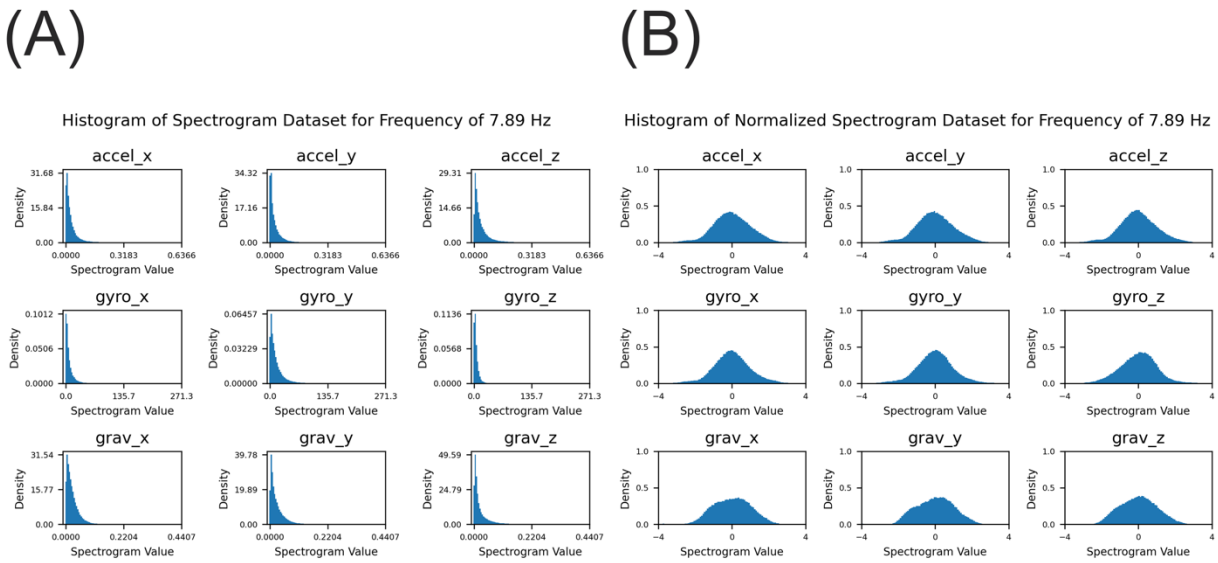



*Figure 28. An example of data conversion using the Box-Cox transformation. (A) The distribution of the amplitude of the spectrum for each sensor measurement. The X-axis represents the amplitude of the signal generated by 9 sensor variables at the selected frequency of 7.89 Hz (I.e., 3-axis acceleration without gravity, 3-axis angular velocity from Gyroscope, and 3-axis vector in the gravity). This frequency is one of the 11 discrete frequencies obtained by dividing the frequency range up to the Nyquist frequency (26.3 Hz) into equal intervals. The Y-axis shows the density distribution of in the value, the integral over the x-axis become 1.0. (B) The distributions of the amplitude of the spectrum after the Box-Cox transformations.*

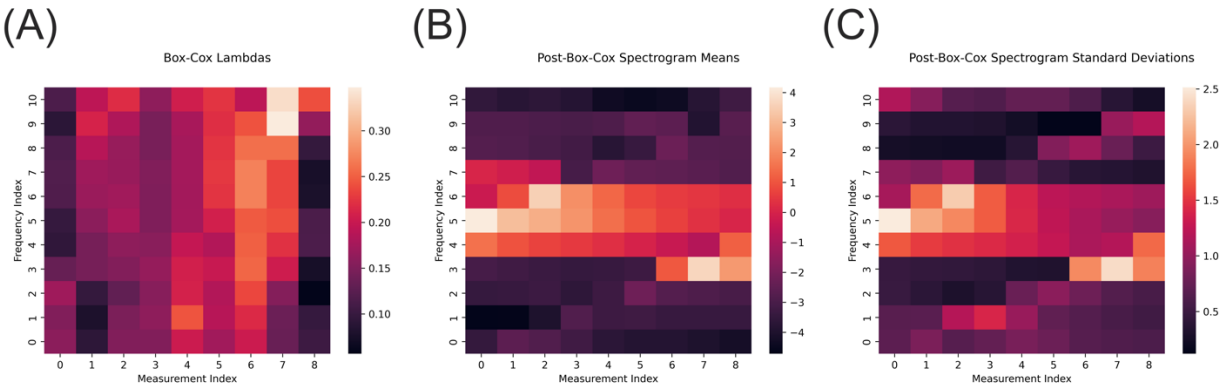



*Figure 29. An example of the heat map of the chosen parameters in Box-Cox transformation. (A) the combination of the lambdas parameters. The lambda parameters were optimally selected to make each distribution close to the normal distribution. The X-axis represents 9 features used in training, and y-axis represents the frequencies determined by the FFT size. (B) The post Box-Cox spectrogram means. (C) The post Box-Cox spectrogram standard deviations.*

*1.2.1. Model Comparisons*

We compared the performances of the CNN to three other machine learning methods: (1) k-nearest neighbor (KNN), (2) Support Vector Machine (SVM), and (3) multi-layer perceptron. A KNN classifies a given data point by finding the majority class of its k nearest neighbors, and we use the Scikit-learn implementation of the KNN. We experimentally set k equal to 11, which yielded the best results for the KNN. An SVM attempts to linearly separate the two classes of points while maximizing the margin between them, and we used the Scikit-learn SGDClassifier implementation of SVM. To speed up the training of the SVM, we applied the Scikit-learn implementation of the Nystroem approximation with RBF kernel to the training data, which reduces the dimension of the data to 100. The perceptron consists of four 100-unit dense layers with the ReLU activation and a dropout rate of 0.5 applied after each layer, as well as a single-unit output layer with sigmoid activation. L2 regularization with a lambda of 0.0005 is applied to the weights and biases of the multi-layer perceptron.

Table 6 displays the selected parameters for each network trained using the Manumeter-Home dataset and Mocap-Lab dataset. For the network trained with the Mocap-Lab dataset, we made adjustments to the parameters, such as modifying the convolutional layer size, incorporating a normalization layer alongside the convolutional and dropout layers to enhance learning stability.

*Table 6. Parameter settings for proposed Networks*

| Data set used for training | The number of Convolutional layers and Dropout layers | The convolution kernel size | The number of filters for the first convolution layer | The normalization layer was used | L2 regularization lambda parameter |
|---|---|---|---|---|---|
| Manumeter-Home Dataset | 8 | (5, 5) | 128 | No | 0.001 |
| Mocap-Lab Dataset (Hand/Arm movement set as actual positive) | 7 | (3, 3) | 100 | Yes | 0.004 |
| Mocap-Lab Dataset (Hand/Arm movement set as actual negative) | 7 | (3, 3) | 100 | Yes | 0.004 |

In this study, we compared the performance of four machine learning methods: K-Nearest Neighbors (KNN), Support Vector Machines (SVM), Multi-layer Perceptron, and Convolutional Neural Networks (CNN). For the KNN, SVM, and Perceptron methods, we tested their performances on (A) the raw samples, where we created one-dimensional vectors using raw sensor IMU signals, and (B) the samples after spectrogram preprocessing and Box-Cox normalization. To use these methods with spectrogram data, we reshaped the spectrograms into one-dimensional vectors by flattening the two-dimensional matrices, allowing them to be used as input for KNN, SVM, and Perceptron. However, the proposed CNN was designed to work specifically with spectrograms, as they provide a two-dimensional representation of the time-varying frequency content of the signals, allowing the CNNs to effectively capture local patterns. As a result, we only tested the custom CNNs on the spectrogram data. Therefore, our

comparison includes a total of 7 combinations of machine learning model and data processing options, the accuracies of which on each test are displayed in Table 2.

To evaluate the performance of our models, we used two different data splitting methods. The first method, called Random 5-Fold Cross Validation (CV) grouped by participants, involved randomly partitioning the dataset into training and testing data without considering the UEFM score of the subjects. This approach aimed to assess the robustness and generalizability of our CNN model when mixing a diverse range of hand impairments due to stroke into the training data. We performed 6 iterations of this random fold process, each time creating a new training and testing data split. The second method, which we refer to as UEFM folds (i.e., LOOCV in the main text), involved splitting the data into training and testing data based on subjects' impairment levels, as determined by their UEFM scores. For instance, we trained HARCS using subjects' data in the range of $30 \leq$ UEFM score $< 66$ when assessing subjects' data in the range of UEFM score $< 30$. This approach allowed us to evaluate the performance of the models across different impairment levels.

*Table 7. Model comparison with accuracy (%) for Random 5-fold CV and UEFM folds (i.e., LOOCV in the main text) based on subjects' impairment levels. The table shows the mean accuracy for Random 5-fold CV, where the dataset is randomly partitioned into training and testing data without considering the UEFM score of the subjects (6 iterations), and the mean accuracy for specific UEFM folds, where the data is split based on subjects' impairment levels.*

| | (A) Spectrograms not used | | | (B) Spectrograms used | | | |
|---|---|---|---|---|---|---|---|
| | KNN | SVM | Perceptron | KNN | SVM | Perceptron | CNN |
| Random 5-Fold CV (6 Iterations, Overall Mean) | 61.82 | 66.87 | 72.35 | 73.86 | 74.54 | 76.46 | 77.19 |
| UEFM Folds | | | | | | | |
| [0, 20) | 61.14 | 62.1 | 70 | 78.38 | 78.1 | 79.81 | 81.05 |
| [20, 30) | 60.71 | 68.8 | 76.97 | 74.94 | 77.27 | 78.46 | 78.69 |
| [30, 40) | 58.46 | 65.92 | 71.94 | 69.64 | 73.92 | 76.3 | 76.68 |
| [40, 50) | 61.77 | 72.8 | 75.23 | 77.43 | 79.66 | 80.56 | 80.61 |
| [50, 60) | 58.99 | 62.36 | 64.31 | 71.97 | 73.38 | 74.59 | 74.44 |
| Mean | 60.21 | 66.4 | 71.69 | 74.47 | 76.47 | 77.94 | 78.29 |

*1.3. Supplementary Material for Chapter 3*

In this supplementary material, we present detailed information on two critical optimizations in our study: (1) Optimization of the Sample Entropy Calculation Parameters, and (2) Optimization of Inactivity Filtering. These optimizations were necessary to ensure a more accurate and unbiased analysis of stroke patients' upper extremity impairment. If not optimized, inactivity filtering may introduce significant bias, and unoptimized sample entropy parameters may hinder our ability to fully benefit from the sample entropy calculation. By addressing these optimizations, we aimed to obtain meaningful statistical features that are related to subjects' clinical scores, which in turn can contribute to a better understanding of arm movement patterns in stroke recovery and potentially the development of more effective therapeutic interventions.

*1.3.1. Optimization of the Sample Entropy Calculation Parameters*

SampEn is a powerful tool for measuring the complexity and predictability of time series data for analyzing the intricate movements and postures of stroke patients' arms, and for quantifying the complexity of physiological measurements, such as EEG and EMG signals . It computes the number of matchings with a template having a length m (i.e., $B^m(\gamma)$), and m+1 (i.e., $B^{m+1}(\gamma)$), and then computes the ratio of matching counts between $B^{m+1}(\gamma)$ and $B^{m+1}(\gamma)$.

To optimize the SampEn calculation for our application, we selected the following parameters: (1) template length (m), (2) tolerance (r), (3) segmentation length (N), (4) sampling rate, and (5) type of sensor signals. Template length (m) represents a length of the template to compare the signal to the rest of the data, while tolerance (r) checks if the

difference between a template and an inspected window is acceptable. Sampling rate is also important, as SampEn may not be good at identifying the complexity of high-resolution sensor measurements due to the nature of the algorithm.

In our analysis, we applied SampEn to (i) the amplitude of acceleration, (ii) the amplitude of angular velocity, and (iii) the tilt angle computed by Equation 3 in the main text. We tested template lengths of m = [2, 3, 4, 5] and tolerances of r = [0.1 – 0.3], and downsampled IMU signals from 52.6 Hz to 26, and 13 Hz, respectively. The parameters were determined based on previous studies and computational complexity considerations.

Our optimization criterion was to maximize the correlation between entropy features and the UEFM score. We then evaluated the best parameter combinations for acceleration, angular velocity, and tilt angle. Figures 30-32 display the parameter optimization for SampEn with acceleration, angular velocity, and tilt angle, respectively. An ordinary least square fit was conducted for each combination, and an $R^2$ was derived from the linear fit. Consequently, different parameters were used for each variable (Table 9).

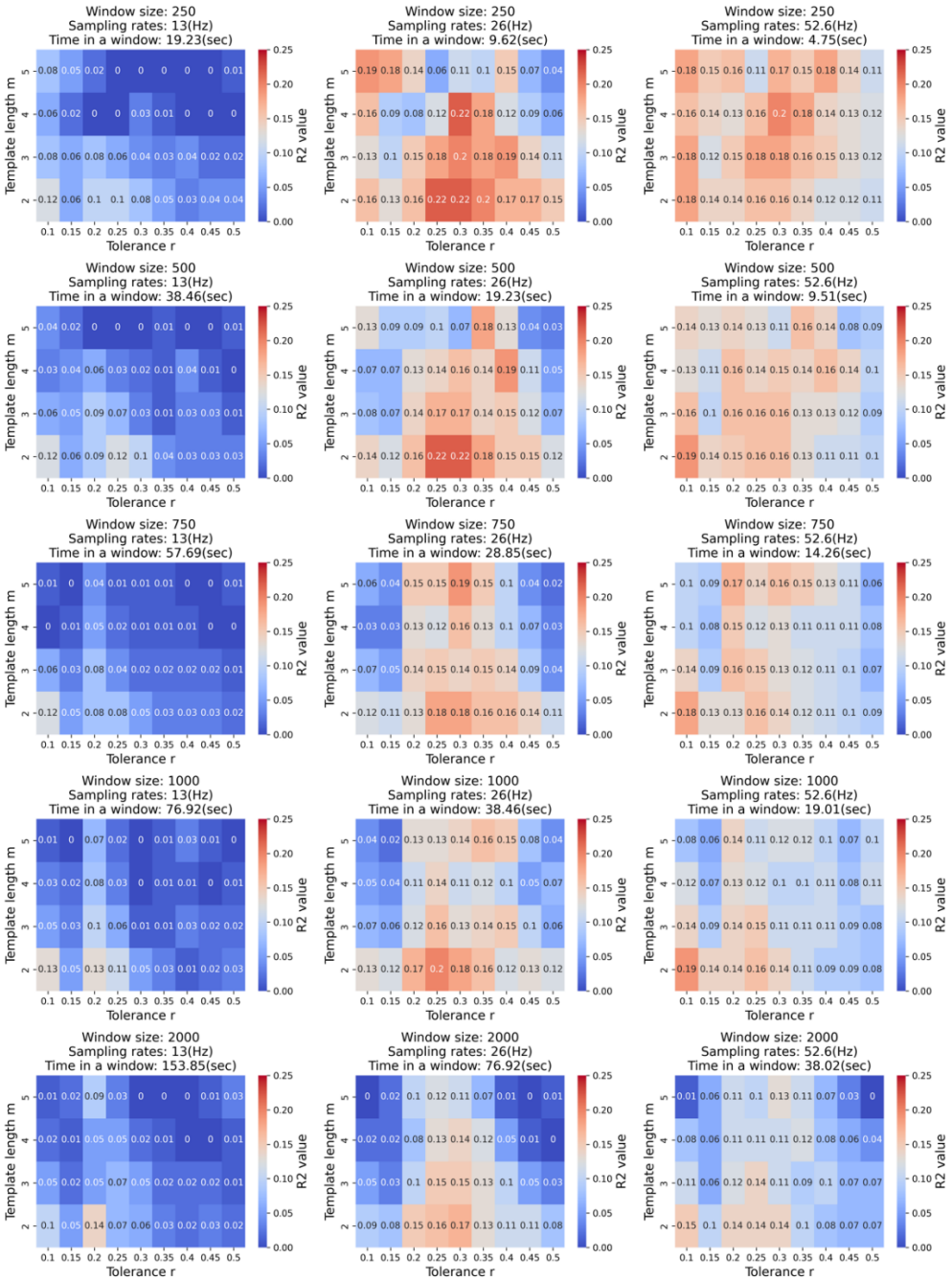



*Figure 30. Parameter optimization of SampEn with acceleration signals.*

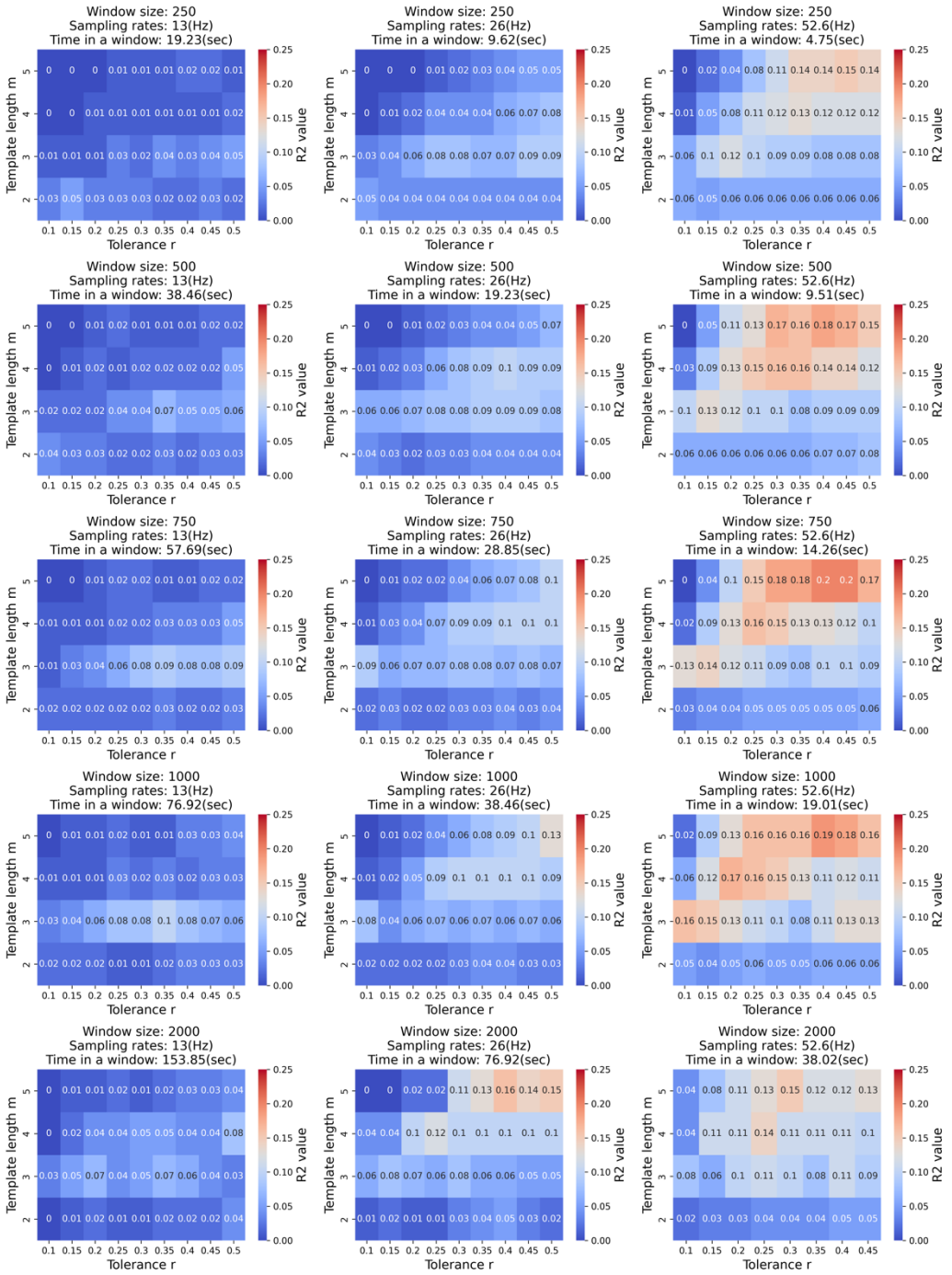



*Figure 31. Parameter optimization of SampEn with angular velocity signals*

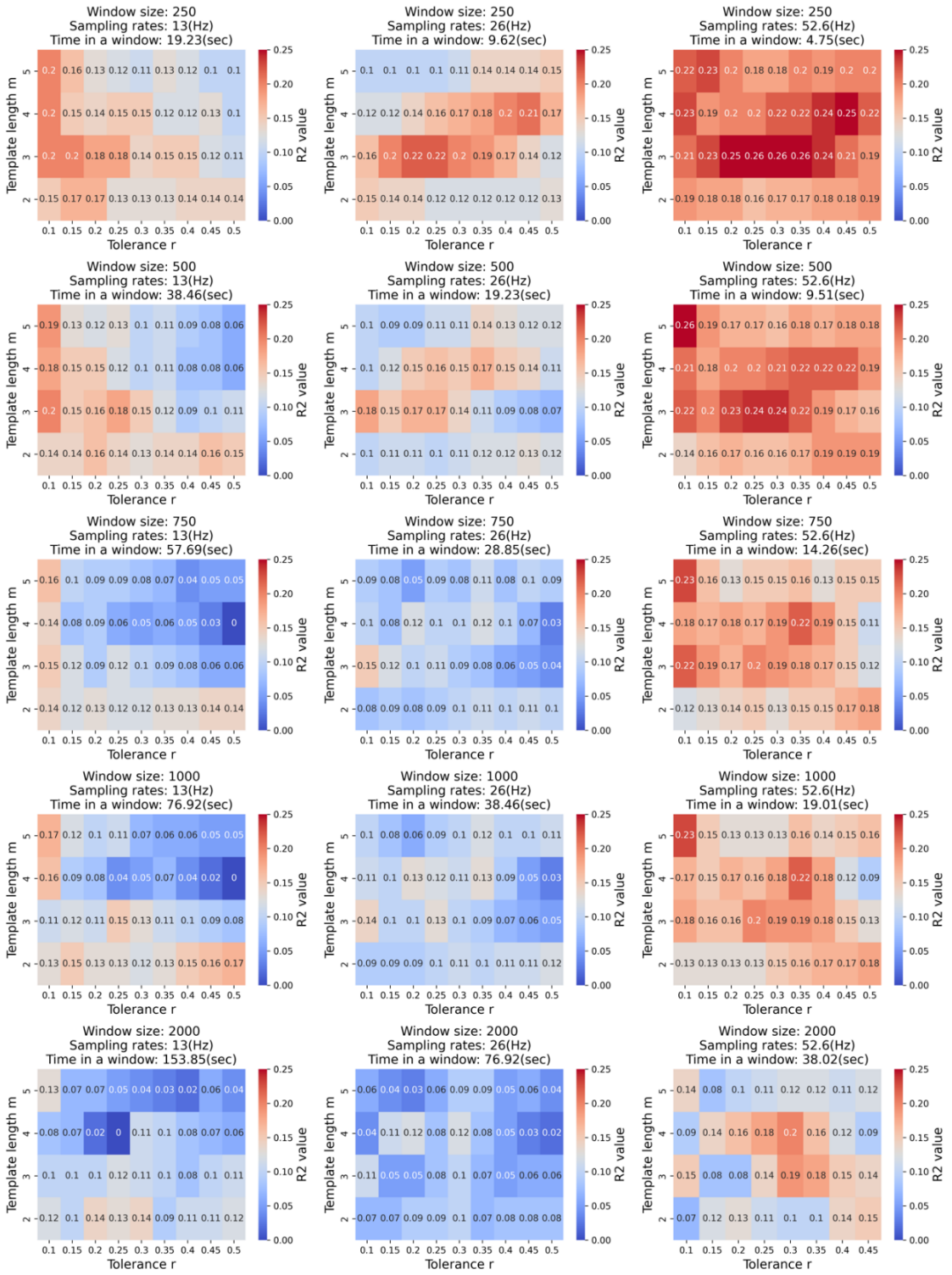



*Figure 32. Parameter optimization of SampEn with tilt angle signals*

*Table 8. Parameter selection for three variables.*

| | Acceleration | Angular velocity | Tilt angle |
|---|---|---|---|
| Template length | 2 | 5 | 3 |
| Tolerance | 0.25 | 0.4 | 0.35 |
| Segmentation length | 250 | 750 | 250 |
| Sampling rate | 26 | 52.6 | 52.6 |

*1.3.2. Optimization of inactivity filtering*

This section describes the optimization process for inactivity filtering in the context of stroke patients' upper extremity (UE) impairment analysis. Inactivity filtering is essential to remove periods of rest or non-movement from the dataset, as these periods may not contribute meaningfully to the analysis. We used the Madgwick filter to subtract the gravity components from acceleration and introduced two measures of arm activity: (1) the instantaneous upper limb use u(t), and (2) the average arm use U(t).

To find the optimal thresholds for inactivity filtering, we conducted a grid search in the range of [0.05, 0.30] for the threshold of u(t) and [0.05, 0.30] for the threshold of U(t). This approach allowed us to identify the best combination of parameters that yielded a statistically significant correlation with the UE Fugl-Meyer (UEFM) score.

Throughout the optimization process, we maintained the same parameters to eliminate inactivity consistently across all participants. Our goal was to make the thresholds as lenient as possible to retain as much data as possible for analysis. However, we imposed the constraint that the chosen thresholds should produce a total inactive time that was as strongly correlated as possible with the impairment level measured by UEFM score.

Our goal was to make the thresholds as lenient as possible to retain as much data as possible for analysis. However, we imposed the constraint that the chosen thresholds should produce a total inactive time that was as strongly correlated as possible with the impairment level measured by UEFM score.

We observed a highly statistically significant difference when the threshold for u(t) = 0.1 and the threshold for U(t) = 0.1 (Figure4A, p=0.01). To retain as much information as possible, we selected this combination as the threshold for inactivity filtering.

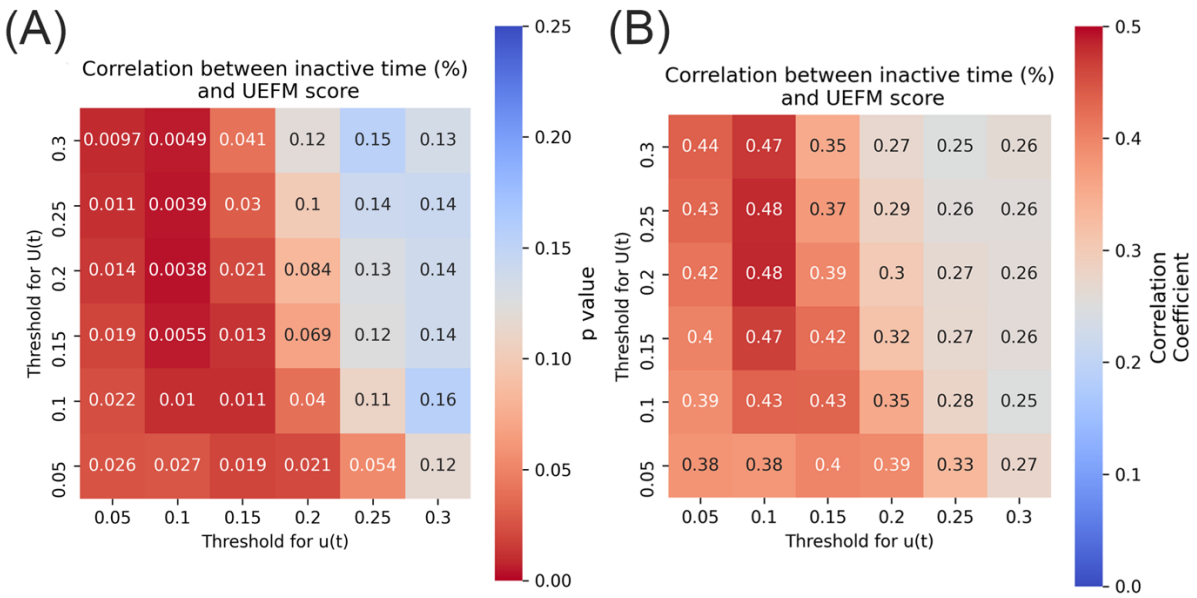



*Figure 33. Optimization of thresholds.*

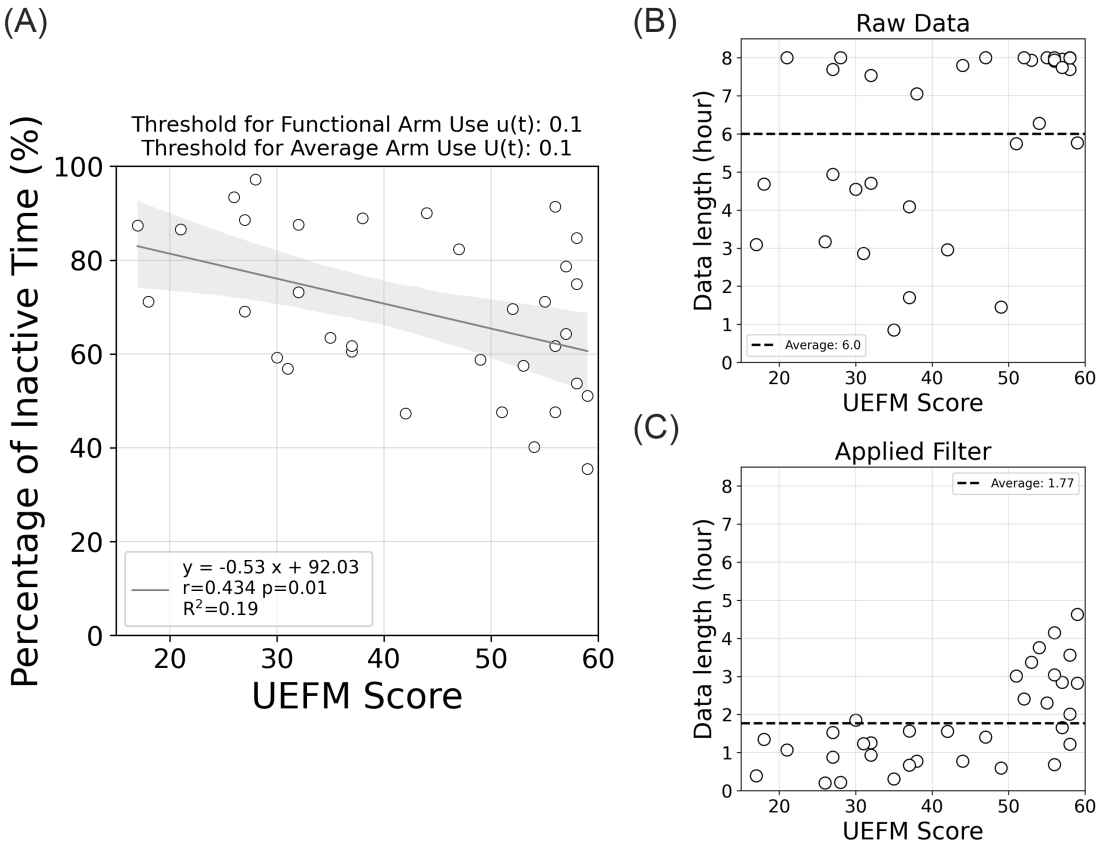



*Figure 34. The best selection of percentage of the inactive time versus the UEFM score with chosen thresholds. (A) Percentage of the inactivity obtained from a black solid line represents a linear regression line (y=-0.53x + 92.03; Pearson correlation, r = 0.43, p=0.01, $R^2$=0.19) and the shaded area represents a confidence interval. (B) The amount of data and UEFM score without filtering. (C) The amount of data after filtering.*

# References


[1] World Health Organization, *World health statistics 2021: monitoring health for the SDGs, sustainable development goals*. World Health Organization, 2021. Accessed: Mar. 07, 2022. [Online]. Available: https://apps.who.int/iris/handle/10665/342703

[2] V. L. Feigin *et al.*, “Global, regional, and national burden of stroke and its risk factors, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019,” *Lancet Neurol.*, vol. 20, no. 10, pp. 795–820, Oct. 2021, doi: 10.1016/S1474-4422(21)00252-0.

[3] G. E. Gresham *et al.*, “Post-Stroke Rehabilitation. Clinical Practice Guideline Number 16; Post-Stroke Rehabilitation: Assessment, Referral, and Patient Management. Quick Reference Guide for Clinicians Number 16; Recovering After a Stroke, Consumer Version. Patient and Family Guide.,” Agency for Health Care Policy and Research, Rockville, MD. Center for Research Dissemination and Liaison., PB95226890, 1995. Accessed: Apr. 23, 2022. [Online]. Available: https://ntrl.ntis.gov/NTRL/dashboard/searchResults/titleDetail/PB95226890.xhtml

[4] E. H. Lo, T. Dalkara, and M. A. Moskowitz, “Mechanisms, challenges and opportunities in stroke,” *Nat. Rev. Neurosci.*, vol. 4, no. 5, Art. no. 5, May 2003, doi: 10.1038/nrn1106.

[5] B. Zaaimi, S. A. Edgley, D. S. Soteropoulos, and S. N. Baker, “Changes in descending motor pathway connectivity after corticospinal tract lesion in macaque monkey,” *Brain*, vol. 135, no. 7, pp. 2277–2289, Jul. 2012, doi: 10.1093/brain/aws115.

[6] M. L. Ingemanson *et al.*, “Neural Correlates of Passive Position Finger Sense After Stroke,” *Neurorehabil. Neural Repair*, vol. 33, no. 9, 2019, doi: 10.1177/1545968319862556.

[7] M. Ghassemi and D. G. Kamper, “The hand after stroke and SCI: Restoration of function with technology,” in *Neurorehabilitation Technology*, D. J. Reinkensmeyer, L. Marchal-Crespo, and V. Dietz, Eds., 3rd ed.Springer, 2022.

[8] D. S. Nichols-Larsen, P. c. Clark, A. Zeringue, A. Greenspan, and S. Blanton, “Factors Influencing Stroke Survivors’ Quality of Life During Subacute Recovery,” *Stroke*, vol. 36, no. 7, pp. 1480–1484, Jul. 2005, doi: 10.1161/01.STR.0000170706.13595.4f.

[9] R. W. Teasell, N. C. Foley, S. K. Bhogal, and M. R. Speechley, “An evidence-based review of stroke rehabilitation,” *Top. Stroke Rehabil.*, vol. 10, no. 1, pp. 29–58, 2003, doi: 10.1310/8YNA-1YHK-YMHB-XTE1.

[10] C. E. Lang, K. J. Waddell, J. Barth, C. L. Holleran, M. J. Strube, and M. D. Bland, “Upper Limb Performance in Daily Life Approaches Plateau Around Three to Six Weeks Post-stroke,” *Neurorehabil. Neural Repair*, vol. 35, no. 10, pp. 903–914, Oct. 2021, doi: 10.1177/15459683211041302.

[11] L. Li, Q. Fu, S. Tyson, N. Preston, and A. Weightman, “A scoping review of design requirements for a home-based upper limb rehabilitation robot for stroke,” *Top. Stroke Rehabil.*, vol. 29, no. 6, pp. 449–463, Aug. 2022, doi: 10.1080/10749357.2021.1943797.

[12] M. Maier, B. R. Ballester, and P. F. M. J. Verschure, “Principles of Neurorehabilitation After Stroke Based on Motor Learning and Brain Plasticity

Mechanisms," *Front. Syst. Neurosci.*, vol. 13, 2019, Accessed: Feb. 09, 2022. [Online]. Available: https://www.frontiersin.org/article/10.3389/fnsys.2019.00074

[13] K. R. Lohse, C. E. Lang, and L. A. Boyd, "Is More Better? Using Metadata to Explore Dose–Response Relationships in Stroke Rehabilitation," *Stroke*, vol. 45, no. 7, pp. 2053–2058, Jul. 2014, doi: 10.1161/STROKEAHA.114.004695.

[14] C. E. Lang *et al.*, "Dose response of task-specific upper limb training in people at least 6 months poststroke: A phase II, single-blind, randomized, controlled trial," *Ann. Neurol.*, vol. 80, no. 3, pp. 342–354, 2016, doi: 10.1002/ana.24734.

[15] C. Winstein, B. Kim, S. Kim, C. Martinez, and N. Schweighofer, "Dosage Matters," *Stroke*, vol. 50, no. 7, pp. 1831–1837, Jul. 2019, doi: 10.1161/STROKEAHA.118.023603.

[16] C. P. Montoya, L. J. Campbell-Hope, K. D. Pemberton, and S. B. Dunnett, "The 'staircase test': a measure of independent forelimb reaching and grasping abilities in rats," *J. Neurosci. Methods*, vol. 36, no. 2, pp. 219–228, Feb. 1991, doi: 10.1016/0165-0270(91)90048-5.

[17] M. S. Jeffers *et al.*, "Does Stroke Rehabilitation Really Matter? Part B: An Algorithm for Prescribing an Effective Intensity of Rehabilitation," *Neurorehabil. Neural Repair*, vol. 32, no. 1, pp. 73–83, Jan. 2018, doi: 10.1177/1545968317753074.

[18] B. R. Ballester, C. Winstein, and N. Schweighofer, "Virtuous and Vicious Cycles of Arm Use and Function Post-stroke," *Front. Neurol.*, vol. 13, 2022, Accessed: May 14, 2022. [Online]. Available: https://www.frontiersin.org/article/10.3389/fneur.2022.804211

[19] C. E. Han, M. A. Arbib, and N. Schweighofer, “Stroke Rehabilitation Reaches a Threshold,” *PLOS Comput. Biol.*, vol. 4, no. 8, p. e1000133, Aug. 2008, doi: 10.1371/journal.pcbi.1000133.

[20] N. Schweighofer, C. E. Han, S. L. Wolf, M. A. Arbib, and C. J. Winstein, “A Functional Threshold for Long-Term Use of Hand and Arm Function Can Be Determined: Predictions From a Computational Model and Supporting Data From the Extremity Constraint-Induced Therapy Evaluation (EXCITE) Trial,” *Phys. Ther.*, vol. 89, no. 12, pp. 1327–1336, Dec. 2009, doi: 10.2522/ptj.20080402.

[21] C. E. Lang, J. R. MacDonald, and C. Gnip, “Counting Repetitions: An Observational Study of Outpatient Therapy for People with Hemiparesis Post-Stroke,” *J. Neurol. Phys. Ther.*, vol. 31, no. 1, pp. 3–10, Mar. 2007, doi: 10.1097/01.NPT.0000260568.31746.34.

[22] D. Rand and J. J. Eng, “Disparity Between Functional Recovery and Daily Use of the Upper and Lower Extremities During Subacute Stroke Rehabilitation,” *Neurorehabil. Neural Repair*, vol. 26, no. 1, pp. 76–84, Jan. 2012, doi: 10.1177/1545968311408918.

[23] C. A. Doman, K. J. Waddell, R. R. Bailey, J. L. Moore, and C. E. Lang, “Changes in Upper-Extremity Functional Capacity and Daily Performance During Outpatient Occupational Therapy for People With Stroke,” *Am. J. Occup. Ther.*, vol. 70, no. 3, pp. 7003290040p1-7003290040p11, Apr. 2016, doi: 10.5014/ajot.2016.020891.

[24] C. P. Adans-Dester, C. E. Lang, D. J. Reinkensmeyer, and P. Bonato, “Wearable Sensors for Stroke Rehabilitation,” p. 44.

[25] C. E. Lang *et al.*, “Improvement in the Capacity for Activity Versus Improvement in Performance of Activity in Daily Life During Outpatient Rehabilitation,” *J. Neurol. Phys. Ther.*, vol. 47, no. 1, p. 16, Jan. 2023, doi: 10.1097/NPT.0000000000000413.

[26] D. J. Gladstone, C. J. Danells, and S. E. Black, “The Fugl-Meyer Assessment of Motor Recovery after Stroke: A Critical Review of Its Measurement Properties,” *Neurorehabil. Neural Repair*, vol. 16, no. 3, pp. 232–240, Sep. 2002, doi: 10.1177/154596802401105171.

[27] S. L. Wolf, P. A. Catlin, M. Ellis, A. L. Archer, B. Morgan, and A. Piacentino, “Assessing Wolf Motor Function Test as Outcome Measure for Research in Patients After Stroke,” *Stroke*, vol. 32, no. 7, pp. 1635–1639, Jul. 2001, doi: 10.1161/01.STR.32.7.1635.

[28] D. M. Morris, G. Uswatte, J. E. Crago, E. W. Cook, and E. Taub, “The reliability of the Wolf Motor Function Test for assessing upper extremity function after stroke,” *Arch. Phys. Med. Rehabil.*, vol. 82, no. 6, pp. 750–755, Jun. 2001, doi: 10.1053/apmr.2001.23183.

[29] R. C. Lyle, “A performance test for assessment of upper limb function in physical rehabilitation treatment and research,” *Int. J. Rehabil. Res.*, vol. 4, no. 4, p. 483, Dec. 1981.

[30] N. Yozbatiran, L. Der-Yeghiaian, and S. C. Cramer, “A Standardized Approach to Performing the Action Research Arm Test,” *Neurorehabil. Neural Repair*, vol. 22, no. 1, pp. 78–90, Jan. 2008, doi: 10.1177/1545968307305353.

[31] M. A. Murphy, C. Willén, and K. S. Sunnerhagen, “Kinematic Variables Quantifying Upper-Extremity Performance After Stroke During Reaching and Drinking From a Glass,” *Neurorehabil. Neural Repair*, vol. 25, no. 1, pp. 71–80, Jan. 2011, doi: 10.1177/1545968310370748.

[32] J. Likitlersuang, E. R. Sumitro, T. Cao, R. J. Visée, S. Kalsi-Ryan, and J. Zariffa, “Egocentric video: a new tool for capturing hand use of individuals with spinal cord injury at home,” *J. NeuroEngineering Rehabil.*, vol. 16, no. 1, p. 83, Jul. 2019, doi: 10.1186/s12984-019-0557-1.

[33] A. I. Cuesta-Vargas, A. Galán-Mercant, and J. M. Williams, “The use of inertial sensors system for human motion analysis,” *Phys. Ther. Rev.*, vol. 15, no. 6, pp. 462–473, Dec. 2010, doi: 10.1179/1743288X11Y.0000000006.

[34] C. P. Adans-Dester, C. E. Lang, D. J. Reinkensmeyer, and P. Bonato, “Wearable Sensors for Stroke Rehabilitation,” in *Neurorehabilitation Technology*, D. J. Reinkensmeyer, L. Marchal-Crespo, and V. Dietz, Eds., Cham: Springer International Publishing, 2022, pp. 467–507. doi: 10.1007/978-3-031-08995-4_21.

[35] “Fitness Tracker Market Size, Share, Growth & Analysis [2028].” https://www.fortunebusinessinsights.com/fitness-tracker-market-103358 (accessed Apr. 27, 2022).

[36] D. M. Bravata *et al.*, “Using Pedometers to Increase Physical Activity and Improve Health: A Systematic Review,” *JAMA*, vol. 298, no. 19, p. 2296, Nov. 2007, doi: 10.1001/jama.298.19.2296.

[37] M. Demers *et al.*, “Wearable technology to capture arm use of stroke survivors in home and community settings: feasibility and insights on motor performance.”

medRxiv, p. 2023.01.25.23284790, Jan. 28, 2023. doi: 10.1101/2023.01.25.23284790.

[38] J. P. O. Held, A. R. Luft, and J. M. Veerbeek, “Encouragement-Induced Real-World Upper Limb Use after Stroke by a Tracking and Feedback Device: A Study Protocol for a Multi-Center, Assessor-Blinded, Randomized Controlled Trial,” *Front. Neurol.*, vol. 9, p. 13, Jan. 2018, doi: 10.3389/fneur.2018.00013.

[39] M. Widmer *et al.*, “Reward During Arm Training Improves Impairment and Activity After Stroke: A Randomized Controlled Trial,” *Neurorehabil. Neural Repair*, p. 15459683211062898, Dec. 2021, doi: 10.1177/15459683211062898.

[40] J. B. Rowe, N. Friedman, M. Bachman, and D. J. Reinkensmeyer, “The Manumeter: A non-obtrusive wearable device for monitoring spontaneous use of the wrist and fingers,” in *2013 IEEE 13th International Conference on Rehabilitation Robotics (ICORR)*, Jun. 2013, pp. 1–6. doi: 10.1109/ICORR.2013.6650397.

[41] N. Friedman, J. B. Rowe, D. J. Reinkensmeyer, and M. Bachman, “The manumeter: a wearable device for monitoring daily use of the wrist and fingers,” *IEEE J. Biomed. Health Inform.*, vol. 18, no. 6, pp. 1804–1812, Nov. 2014, doi: 10.1109/JBHI.2014.2329841.

[42] J. E. Sasaki, D. John, and P. S. Freedson, “Validation and comparison of ActiGraph activity monitors,” *J. Sci. Med. Sport*, vol. 14, no. 5, pp. 411–416, Sep. 2011, doi: 10.1016/j.jsams.2011.04.003.

[43] L. Brennan, E. Dorronzoro Zubiete, and B. Caulfield, “Feedback Design in Targeted Exercise Digital Biofeedback Systems for Home Rehabilitation: A Scoping Review,” *Sensors*, vol. 20, no. 1, p. 181, Dec. 2019, doi: 10.3390/s20010181.

[44] L. Oppici, A. Dix, and S. Narciss, “When is knowledge of performance (KP) superior to knowledge of results (KR) in promoting motor skill learning? A systematic review,” *Int. Rev. Sport Exerc. Psychol.*, vol. 0, no. 0, pp. 1–25, Oct. 2021, doi: 10.1080/1750984X.2021.1986849.

[45] Q. Wang, P. Markopoulos, B. Yu, W. Chen, and A. Timmermans, “Interactive wearable systems for upper body rehabilitation: a systematic review,” *J. NeuroEngineering Rehabil.*, vol. 14, no. 1, p. 20, Mar. 2017, doi: 10.1186/s12984-017-0229-y.

[46] R. A. Schmidt, T. D. Lee, C. Winstein, G. Wulf, and H. N. Zelaznik, *Motor control and learning: A behavioral emphasis*. Human kinetics, 2018.

[47] G. Uswatte, C. Giuliani, C. Winstein, A. Zeringue, L. Hobbs, and S. L. Wolf, “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.*, vol. 87, no. 10, pp. 1340–1345, Oct. 2006, doi: 10.1016/j.apmr.2006.06.006.

[48] D. Schwerz de Lucena, J. Rowe, V. Chan, and D. J. Reinkensmeyer, “Magnetically Counting Hand Movements: Validation of a Calibration-Free Algorithm and Application to Testing the Threshold Hypothesis of Real-World Hand Use after Stroke,” *Sensors*, vol. 21, no. 4, Art. no. 4, Jan. 2021, doi: 10.3390/s21041502.

[49] X. Liu, S. Rajan, N. Ramasarma, P. Bonato, and S. I. Lee, “The Use of a Finger-Worn Accelerometer for Monitoring of Hand Use in Ambulatory Settings,” *IEEE J. Biomed. Health Inform.*, vol. 23, no. 2, pp. 599–606, Mar. 2019, doi: 10.1109/JBHI.2018.2821136.

[50] Y. Ling, L. P. T. Meer, Z. Yumak, and R. C. Veltkamp, “Usability Test of Exercise Games Designed for Rehabilitation of Elderly Patients After Hip Replacement Surgery: Pilot Study,” *JMIR Serious Games*, vol. 5, no. 4, p. e7969, Oct. 2017, doi: 10.2196/games.7969.

[51] G. Spina, G. Huang, A. Vaes, M. Spruit, and O. Amft, “COPDTrainer: a smartphone-based motion rehabilitation training system with real-time acoustic feedback,” in *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*, in UbiComp ’13. New York, NY, USA: Association for Computing Machinery, Sep. 2013, pp. 597–606. doi: 10.1145/2493432.2493454.

[52] A. B. Vallbo and R. S. Johansson, “Properties of cutaneous mechanoreceptors in the human hand related to touch sensation,” *Hum. Neurobiol.*, vol. 3, no. 1, pp. 3–14, 1984.

[53] C. E. Seim, S. L. Wolf, and T. E. Starner, “Wearable vibrotactile stimulation for upper extremity rehabilitation in chronic stroke: clinical feasibility trial using the VTS Glove,” *J. NeuroEngineering Rehabil.*, vol. 18, no. 1, p. 14, Jan. 2021, doi: 10.1186/s12984-021-00813-7.

[54] C. E. Seim, B. Ritter, T. E. Starner, K. Flavin, M. G. Lansberg, and A. M. Okamura, “Design of a Wearable Vibrotactile Stimulation Device for Individuals With Upper-Limb Hemiparesis and Spasticity,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 1277–1287, 2022, doi: 10.1109/TNSRE.2022.3174808.

[55] M. V. Perrotta, T. Asgeirsdottir, and D. M. Eagleman, “Deciphering Sounds Through Patterns of Vibration on the Skin,” *Neuroscience*, vol. 458, pp. 77–86, Mar. 2021, doi: 10.1016/j.neuroscience.2021.01.008.

[56] W. X. J. Wei, K. N. K. Fong, R. C. K. Chung, H. K. Y. Cheung, and E. S. L. Chow, “‘Remind-to-Move’ for Promoting Upper Extremity Recovery Using Wearable Devices in Subacute Stroke: A Multi-Center Randomized Controlled Study,” *IEEE Trans. Neural Syst. Rehabil. Eng. Publ. IEEE Eng. Med. Biol. Soc.*, vol. 27, no. 1, pp. 51–59, Jan. 2019, doi: 10.1109/TNSRE.2018.2882235.

[57] Z. Bai and K. N. K. Fong, “‘Remind-to-Move’ Treatment Enhanced Activation of the Primary Motor Cortex in Patients with Stroke,” *Brain Topogr.*, vol. 33, no. 2, pp. 275–283, Mar. 2020, doi: 10.1007/s10548-020-00756-7.

[58] S. R. Egglestone *et al.*, “A design framework for a home-based stroke rehabilitation system: Identifying the key components,” in *2009 3rd International Conference on Pervasive Computing Technologies for Healthcare*, Apr. 2009, pp. 1–8. doi: 10.4108/ICST.PERVASIVEHEALTH2009.6049.

[59] N. Nasr *et al.*, “The experience of living with stroke and using technology: opportunities to engage and co-design with end users,” *Disabil. Rehabil. Assist. Technol.*, vol. 11, no. 8, pp. 653–660, Nov. 2016, doi: 10.3109/17483107.2015.1036469.

[60] S. Balasubramanian, A. Melendez-Calderon, and E. Burdet, “A Robust and Sensitive Metric for Quantifying Movement Smoothness,” *IEEE Trans. Biomed. Eng.*, vol. 59, no. 8, pp. 2126–2136, Aug. 2012, doi: 10.1109/TBME.2011.2179545.

[61] B. Rohrer *et al.*, “Movement smoothness changes during stroke recovery,” *J. Neurosci. Off. J. Soc. Neurosci.*, vol. 22, no. 18, pp. 8297–8304, Sep. 2002.

[62] S. Balasubramanian, A. Melendez-Calderon, A. Roby-Brami, and E. Burdet, “On the analysis of movement smoothness,” *J. NeuroEngineering Rehabil.*, vol. 12, no. 1, p. 112, Dec. 2015, doi: 10.1186/s12984-015-0090-9.

[63] S. L. DeJong, S. Y. Schaefer, and C. E. Lang, “Need for speed: better movement quality during faster task performance after stroke,” *Neurorehabil. Neural Repair*, vol. 26, no. 4, pp. 362–373, May 2012, doi: 10.1177/1545968311425926.

[64] D. G. Kamper, A. N. McKenna-Cole, L. E. Kahn, and D. J. Reinkensmeyer, “Alterations in reaching after stroke and their relation to movement direction and impairment severity,” *Arch. Phys. Med. Rehabil.*, vol. 83, no. 5, 2002, doi: 10.1053/apmr.2002.32446.

[65] M. C. Cirstea and M. F. Levin, “Compensatory strategies for reaching in stroke,” *Brain*, vol. 123, no. 5, pp. 940–953, 2000.

[66] J. W. Krakauer *et al.*, “Comparing a Novel Neuroanimation Experience to Conventional Therapy for High-Dose Intensive Upper-Limb Training in Subacute Stroke: The SMARTS2 Randomized Trial,” *Neurorehabil. Neural Repair*, vol. 35, no. 5, pp. 393–405, May 2021, doi: 10.1177/15459683211000730.

[67] I. S. Howard, J. N. Ingram, K. P. Körding, and D. M. Wolpert, “Statistics of Natural Movements Are Reflected in Motor Errors,” *J. Neurophysiol.*, vol. 102, no. 3, pp. 1902–1910, Sep. 2009, doi: 10.1152/jn.00013.2009.

[68] S. Albawi, T. A. Mohammed, and S. Al-Zawi, “Understanding of a convolutional neural network,” in *2017 International Conference on Engineering and Technology (ICET)*, Aug. 2017, pp. 1–6. doi: 10.1109/ICEngTechnol.2017.8308186.

[69] M. Lloyd, M. MacDonald, and C. Lord, “Motor skills of toddlers with autism spectrum disorders,” *Autism*, vol. 17, no. 2, pp. 133–146, Mar. 2013, doi: 10.1177/1362361311402230.

[70] D. Schwerz de Lucena, J. B. Rowe, S. Okita, V. Chan, S. C. Cramer, and D. J. Reinkensmeyer, “Providing Real-Time Wearable Feedback to Increase Hand Use after Stroke: A Randomized, Controlled Trial,” *Sensors*, vol. 22, no. 18, Art. no. 18, Jan. 2022, doi: 10.3390/s22186938.

[71] K. D. Anderson, “Targeting Recovery: Priorities of the Spinal Cord-Injured Population,” *J. Neurotrauma*, vol. 21, no. 10, pp. 1371–1383, Oct. 2004, doi: 10.1089/neu.2004.21.1371.

[72] M. Hagberg, H. Morgenstern, and M. Kelsh, “Impact of occupations and job tasks on the prevalence of carpal tunnel syndrome.,” *Scand. J. Work. Environ. Health*, vol. 18, no. 6, pp. 337–345, 1992, doi: 10.5271/sjweh.1564.

[73] Y. Kim *et al.*, “Towards the Design of a Ring Sensor-based mHealth System to Achieve Optimal Motor Function in Stroke Survivors,” *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 3, no. 4, p. 138:1-138:26, Sep. 2020, doi: 10.1145/3369817.

[74] C. Harrison, D. Tan, and D. Morris, “Skinput: appropriating the body as an input surface,” in *Proceedings of the SIGCHI Conference on Human Factors in*

*Computing Systems*, in CHI '10. New York, NY, USA: Association for Computing Machinery, Apr. 2010, pp. 453–462. doi: 10.1145/1753326.1753394.

[75] O. Banos, J.-M. Galvez, M. Damas, H. Pomares, and I. Rojas, "Window Size Impact in Human Activity Recognition," *Sensors*, vol. 14, no. 4, Art. no. 4, Apr. 2014, doi: 10.3390/s140406474.

[76] "SciPy 1.0: fundamental algorithms for scientific computing in Python | Nature Methods." https://www.nature.com/articles/s41592-019-0686-2?report=reader (accessed Apr. 02, 2022).

[77] D. Schwerz de Lucena, "New Technologies for On-Demand Hand Rehabilitation in the Living Environment after Neurologic Injury," UC Irvine, 2019. Accessed: Feb. 08, 2022. [Online]. Available: https://escholarship.org/uc/item/42423813

[78] S. O. H. Madgwick, "An efficient orientation filter for inertial and inertial/magnetic sensor arrays," p. 32.

[79] M. Bicego and S. Baldo, "Properties of the Box–Cox transformation for pattern classification," *Neurocomputing*, vol. 218, pp. 390–400, Dec. 2016, doi: 10.1016/j.neucom.2016.08.081.

[80] A. Cheddad, "On Box-Cox Transformation for Image Normality and Pattern Classification," *IEEE Access*, vol. 8, pp. 154975–154983, 2020, doi: 10.1109/ACCESS.2020.3018874.

[81] A. C. Atkinson, M. Riani, and A. Corbellini, "The Box–Cox Transformation: Review and Extensions," *Stat. Sci.*, vol. 36, no. 2, pp. 239–255, May 2021, doi: 10.1214/20-STS778.

[82] R. M. Sakia, “The Box-Cox Transformation Technique: A Review,” *J. R. Stat. Soc. Ser. Stat.*, vol. 41, no. 2, pp. 169–178, 1992, doi: 10.2307/2348250.

[83] Q. Zhou *et al.*, “Cough Recognition Based on Mel-Spectrogram and Convolutional Neural Network,” *Front. Robot. AI*, vol. 8, 2021, Accessed: Apr. 09, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/frobt.2021.580080

[84] P. Khunarsal, C. Lursinsap, and T. Raicharoen, “Singing voice recognition based on matching of spectrogram pattern,” in *2009 International Joint Conference on Neural Networks*, Jun. 2009, pp. 1595–1599. doi: 10.1109/IJCNN.2009.5179014.

[85] O. Steven Eyobu and D. S. Han, “Feature Representation and Data Augmentation for Human Activity Classification Based on Wearable IMU Sensor Data Using a Deep LSTM Neural Network,” *Sensors*, vol. 18, no. 9, Art. no. 9, Sep. 2018, doi: 10.3390/s18092892.

[86] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: A Simple Way to Prevent Neural Networks from Overfitting,” *J. Mach. Learn. Res.*, vol. 15, no. 56, pp. 1929–1958, 2014.

[87] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” *ArXiv14126980 Cs*, Jan. 2017, Accessed: Mar. 14, 2022. [Online]. Available: http://arxiv.org/abs/1412.6980

[88] M. Abadi *et al.*, “TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems.” arXiv, Mar. 16, 2016. doi: 10.48550/arXiv.1603.04467.

[89] F. Chollet and others, “Keras,” 2015. https://github.com/fchollet/keras

[90] D. Biswas *et al.*, “Recognizing upper limb movements with wrist worn inertial sensors using k-means clustering classification,” *Hum. Mov. Sci.*, vol. 40, pp. 59–76, Apr. 2015, doi: 10.1016/j.humov.2014.11.013.

[91] M. Panwar *et al.*, “Rehab-Net: Deep Learning Framework for Arm Movement Classification Using Wearable Sensors for Stroke Rehabilitation,” *IEEE Trans. Biomed. Eng.*, vol. 66, no. 11, pp. 3026–3037, Nov. 2019, doi: 10.1109/TBME.2019.2899927.

[92] F. Pedregosa *et al.*, “Scikit-learn: Machine learning in Python,” *J. Mach. Learn. Res.*, vol. 12, no. Oct, pp. 2825–2830, 2011.

[93] A. Tharwat, “Classification assessment methods,” *Appl. Comput. Inform.*, vol. 17, no. 1, pp. 168–192, Jan. 2020, doi: 10.1016/j.aci.2018.08.003.

[94] J. A. Hanley and B. J. McNeil, “The meaning and use of the area under a receiver operating characteristic (ROC) curve.,” *Radiology*, vol. 143, no. 1, pp. 29–36, Apr. 1982, doi: 10.1148/radiology.143.1.7063747.

[95] E. J. Woytowicz *et al.*, “Determining Levels of Upper Extremity Movement Impairment by Applying a Cluster Analysis to the Fugl-Meyer Assessment of the Upper Extremity in Chronic Stroke,” *Arch. Phys. Med. Rehabil.*, vol. 98, no. 3, pp. 456–462, Mar. 2017, doi: 10.1016/j.apmr.2016.06.023.

[96] Y. Feito, D. R. Bassett, and D. L. Thompson, “Evaluation of Activity Monitors in Controlled and Free-Living Environments,” *Med. Sci. Sports Exerc.*, vol. 44, no. 4, pp. 733–741, Apr. 2012, doi: 10.1249/MSS.0b013e3182351913.

[97] D. Fuller *et al.*, “Reliability and Validity of Commercially Available Wearable Devices for Measuring Steps, Energy Expenditure, and Heart Rate: Systematic

Review," *JMIR MHealth UHealth*, vol. 8, no. 9, p. e18694, Sep. 2020, doi: 10.2196/18694.

[98] C. Jayaraman, C. K. Mummidisetty, A. Mannix-Slobig, L. McGee Koch, and A. Jayaraman, "Variables influencing wearable sensor outcome estimates in individuals with stroke and incomplete spinal cord injury: a pilot investigation validating two research grade sensors," *J. NeuroEngineering Rehabil.*, vol. 15, no. 1, p. 19, Mar. 2018, doi: 10.1186/s12984-018-0358-y.

[99] J. Zariffa and M. R. Popovic, "Hand contour detection in wearable camera video using an adaptive histogram region of interest," *J. NeuroEngineering Rehabil.*, vol. 10, no. 1, p. 114, Dec. 2013, doi: 10.1186/1743-0003-10-114.

[100] D.-H. Kim *et al.*, "Epidermal Electronics," *Science*, vol. 333, no. 6044, pp. 838–843, Aug. 2011, doi: 10.1126/science.1206157.

[101] K. K. Kim *et al.*, "A deep-learned skin sensor decoding the epicentral human motions," *Nat. Commun.*, vol. 11, no. 1, Art. no. 1, May 2020, doi: 10.1038/s41467-020-16040-y.

[102] E. S. Lawrence *et al.*, "Estimates of the prevalence of acute stroke impairments and disability in a multiethnic population," *Stroke*, vol. 32, no. 6, pp. 1279–1284, Jun. 2001, doi: 10.1161/01.str.32.6.1279.

[103] K. ANDREWS and J. STEWARD, "SROKE RECOVERY: HE CAN BUT DOES HE?," *Rheumatology*, vol. 18, no. 1, pp. 43–48, Feb. 1979, doi: 10.1093/rheumatology/18.1.43.

[104] Adans-Dester, Catherine P., Lang, C.E., Reinkensmeyer, D. J., and Bonato, P., "Wearable Sensors for Stroke Rehabilitation," in *Neurorehabilitation Technology, 3rd Edition*, Springer, 2022.

[105] M. Demers *et al.*, "Wearable technology to capture arm use of stroke survivors in home and community settings: feasibility and insights on motor performance," *medRxiv*, p. 2023.01.25.23284790, Jan. 2023, doi: 10.1101/2023.01.25.23284790.

[106] E. Taub, G. Uswatte, V. W. Mark, and D. M. Morris, "The learned nonuse phenomenon: A implications for rehabilitation," *Eur. MEDICOPHYSICA*, vol. 42, no. 3, p. 15, 2006.

[107] K. Peek, M. Carey, L. Mackenzie, and R. Sanson-Fisher, "Patient adherence to an exercise program for chronic low back pain measured by patient-report, physiotherapist-perception and observational data," *Physiother. Theory Pract.*, vol. 35, no. 12, pp. 1304–1313, Dec. 2019, doi: 10.1080/09593985.2018.1474402.

[108] D. J. Reinkensmeyer, *Neurorehabilitation Technology*, 3rd ed. Springer:, 2022. Accessed: Oct. 10, 2022. [Online]. Available: https://link.springer.com/book/10.1007/978-1-4471-2277-7

[109] C. E. Lang *et al.*, "Improvement in the Capacity for Activity Versus Improvement in Performance of Activity in Daily Life During Outpatient Rehabilitation," *J. Neurol. Phys. Ther. JNPT*, vol. 47, no. 1, pp. 16–25, Jan. 2023, doi: 10.1097/NPT.0000000000000413.

[110] L. De Wit *et al.*, "Use of time by stroke patients: a comparison of four European rehabilitation centers," *Stroke*, vol. 36, no. 9, pp. 1977–1983, Sep. 2005, doi: 10.1161/01.STR.0000177871.59003.e3.

[111] P. J. A. Nicolson, R. S. Hinman, T. V. Wrigley, P. W. Stratford, and K. L. Bennell, “Self-reported Home Exercise Adherence: A Validity and Reliability Study Using Concealed Accelerometers,” *J. Orthop. Sports Phys. Ther.*, vol. 48, no. 12, pp. 943–950, Dec. 2018, doi: 10.2519/jospt.2018.8275.

[112] M. Whitford, E. Schearer, and M. Rowlett, “Effects of in home high dose accelerometer-based feedback on perceived and actual use in participants chronic post-stroke,” *Physiother. Theory Pract.*, vol. 36, no. 7, pp. 799–809, Jul. 2020, doi: 10.1080/09593985.2018.1493759.

[113] R. H. Da-Silva *et al.*, “Prompting arm activity after stroke: A clinical proof of concept study of wrist-worn accelerometers with a vibrating alert function,” *J. Rehabil. Assist. Technol. Eng.*, vol. 5, p. 2055668318761524, Dec. 2018, doi: 10.1177/2055668318761524.

[114] P. Gulde and J. Hermsdörfer, “Smoothness Metrics in Complex Movement Tasks,” *Front. Neurol.*, vol. 9, 2018, Accessed: Jun. 20, 2022. [Online]. Available: https://www.frontiersin.org/article/10.3389/fneur.2018.00615

[115] D. G. Kamper, A. N. McKenna-Cole, L. E. Kahn, and D. J. Reinkensmeyer, “Alterations in reaching after stroke and their relation to movement direction and impairment severity,” *Arch. Phys. Med. Rehabil.*, vol. 83, no. 5, pp. 702–707, May 2002, doi: 10.1053/apmr.2002.32446.

[116] A. Doussoulin *et al.*, “Prevalence of Spasticity and Postural Patterns in the Upper Extremity Post Stroke,” *J. Stroke Cerebrovasc. Dis.*, vol. 29, no. 11, p. 105253, Nov. 2020, doi: 10.1016/j.jstrokecerebrovasdis.2020.105253.

[117] V. Mathiowetz, G. Volland, N. Kashman, and K. Weber, “Adult norms for the Box and Block Test of manual dexterity,” *Am J Occup Ther*, vol. 39, no. 6, pp. 386–391, 1985.

[118] C. B. Lundquist and T. Maribo, “The Fugl-Meyer assessment of the upper extremity: reliability, responsiveness and validity of the Danish version,” *Disabil. Rehabil.*, vol. 39, no. 9, pp. 934–939, May 2017, doi: 10.3109/09638288.2016.1163422.

[119] A. David, T. Subash, S. K. M. Varadhan, A. Melendez-Calderon, and S. Balasubramanian, “A Framework for Sensor-Based Assessment of Upper-Limb Functioning in Hemiparesis,” *Front. Hum. Neurosci.*, vol. 15, 2021, Accessed: Mar. 01, 2022. [Online]. Available: https://www.frontiersin.org/article/10.3389/fnhum.2021.667509

[120] A. David *et al.*, “Quantification of the relative arm use in patients with hemiparesis using inertial measurement units,” *J. Rehabil. Assist. Technol. Eng.*, vol. 8, p. 20556683211019696, Jan. 2021, doi: 10.1177/20556683211019694.

[121] K. Leuenberger, R. Gonzenbach, S. Wachter, A. Luft, and R. Gassert, “A method to qualitatively assess arm use in stroke survivors in the home environment,” *Med. Biol. Eng. Comput.*, vol. 55, no. 1, pp. 141–150, Jan. 2017, doi: 10.1007/s11517-016-1496-7.

[122] “AN-1057: Using an Accelerometer for Inclination Sensing | Analog Devices.” https://www.analog.com/en/app-notes/an-1057.html (accessed Apr. 21, 2022).

[123] M. Pedley, “Tilt Sensing Using a Three-Axis Accelerometer,” *Free. Semicond. Appl. Note 1*, 2013.

[124] J. S. Richman and J. R. Moorman, “Physiological time-series analysis using approximate entropy and sample entropy,” *Am. J. Physiol.-Heart Circ. Physiol.*, vol. 278, no. 6, pp. H2039–H2049, Jun. 2000, doi: 10.1152/ajpheart.2000.278.6.H2039.

[125] J. S. Richman, D. E. Lake, and J. R. Moorman, “Sample Entropy,” in *Methods in Enzymology*, in Numerical Computer Methods, Part E, vol. 384. Academic Press, 2004, pp. 172–184. doi: 10.1016/S0076-6879(04)84011-4.

[126] L. Montesinos, R. Castaldo, and L. Pecchia, “On the use of approximate entropy and sample entropy with centre of pressure time-series,” *J. NeuroEngineering Rehabil.*, vol. 15, no. 1, p. 116, Dec. 2018, doi: 10.1186/s12984-018-0465-9.

[127] Y. Kaur *et al.*, “The reliability and psychometric structure of Multi-Scale Entropy measured from EEG signals at rest and during face and object recognition tasks,” *J. Neurosci. Methods*, vol. 326, p. 108343, Oct. 2019, doi: 10.1016/j.jneumeth.2019.108343.

[128] X. Zhang and P. Zhou, “Sample entropy analysis of surface EMG for improved muscle activity onset detection against spurious background spikes,” *J. Electromyogr. Kinesiol. Off. J. Int. Soc. Electrophysiol. Kinesiol.*, vol. 22, no. 6, pp. 901–907, Dec. 2012, doi: 10.1016/j.jelekin.2012.06.005.

[129] J. D. McCamley, W. Denton, A. Arnold, P. C. Raffalt, and J. M. Yentes, “On the Calculation of Sample Entropy Using Continuous and Discrete Human Gait Data,” *Entropy*, vol. 20, no. 10, Art. no. 10, Oct. 2018, doi: 10.3390/e20100764.

[130] M. L. Woodbury, C. A. Velozo, L. G. Richards, and P. W. Duncan, “Rasch Analysis Staging Methodology to Classify Upper Extremity Movement Impairment

After Stroke," *Arch. Phys. Med. Rehabil.*, vol. 94, no. 8, pp. 1527–1533, Aug. 2013, doi: 10.1016/j.apmr.2013.03.007.

[131] T. Platz, C. Pinkowski, F. van Wijck, I.-H. Kim, P. di Bella, and G. Johnson, "Reliability and validity of arm function assessment with standardized guidelines for the Fugl-Meyer Test, Action Research Arm Test and Box and Block Test: a multicentre study," *Clin. Rehabil.*, vol. 19, no. 4, pp. 404–411, Jun. 2005, doi: 10.1191/0269215505cr832oa.

[132] G. M. Sullivan and R. Feinn, "Using Effect Size—or Why the P Value Is Not Enough," *J. Grad. Med. Educ.*, vol. 4, no. 3, pp. 279–282, Sep. 2012, doi: 10.4300/JGME-D-12-00156.1.

[133] G. Heinze, C. Wallisch, and D. Dunkler, "Variable selection – A review and recommendations for the practicing statistician," *Biom. J.*, vol. 60, no. 3, pp. 431–449, 2018, doi: 10.1002/bimj.201700067.

[134] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Autom. Control*, vol. 19, no. 6, pp. 716–723, Dec. 1974, doi: 10.1109/TAC.1974.1100705.

[135] H. Akaike, "Information Theory and an Extension of the Maximum Likelihood Principle," in *Selected Papers of Hirotugu Akaike*, E. Parzen, K. Tanabe, and G. Kitagawa, Eds., in Springer Series in Statistics. New York, NY: Springer, 1998, pp. 199–213. doi: 10.1007/978-1-4612-1694-0_15.

[136] H. C. V. Houwelingen and W. Sauerbrei, "Cross-Validation, Shrinkage and Variable Selection in Linear Regression Revisited," *Open J. Stat.*, vol. 03, no. 02, pp. 79–102, 2013, doi: 10.4236/ojs.2013.32011.

[137] C. Lindsey and S. Sheather, “Variable Selection in Linear Regression,” *Stata J.*, 2010.

[138] M. Snipes and D. C. Taylor, “Model selection and Akaike Information Criteria: An example from wine ratings and prices,” *Wine Econ. Policy*, vol. 3, no. 1, pp. 3–9, Jun. 2014, doi: 10.1016/j.wep.2014.03.001.

[139] K. P. Burnham and D. R. Anderson, “Multimodel Inference: Understanding AIC and BIC in Model Selection,” *Sociol. Methods Res.*, vol. 33, no. 2, pp. 261–304, Nov. 2004, doi: 10.1177/0049124104268644.

[140] B. S. Chissom, “Interpretation of the Kurtosis Statistic,” *Am. Stat.*, vol. 24, no. 4, pp. 19–22, Oct. 1970, doi: 10.1080/00031305.1970.10477202.

[141] R. M. O’brien, “A Caution Regarding Rules of Thumb for Variance Inflation Factors,” *Qual. Quant.*, vol. 41, no. 5, pp. 673–690, Oct. 2007, doi: 10.1007/s11135-006-9018-6.

[142] H. Hefter, W. H. Jost, A. Reissig, B. Zakine, A. M. Bakheit, and J. Wissel, “Classification of posture in poststroke upper limb spasticity: a potential decision tool for botulinum toxin A treatment?,” *Int. J. Rehabil. Res. Int. Z. Rehabil. Rev. Int. Rech. Readaptation*, vol. 35, no. 3, pp. 227–233, Sep. 2012, doi: 10.1097/MRR.0b013e328353e3d4.

[143] W. H. Jost, H. Hefter, A. Reissig, K. Kollewe, and J. Wissel, “Efficacy and safety of botulinum toxin type A (Dysport) for the treatment of post-stroke arm spasticity: Results of the German–Austrian open-label post-marketing surveillance prospective study,” *J. Neurol. Sci.*, vol. 337, no. 1, pp. 86–90, Feb. 2014, doi: 10.1016/j.jns.2013.11.022.

[144] D. J. Reinkensmeyer, A. McKenna Cole, L. E. Kahn, and D. G. Kamper, "Directional control of reaching is preserved following mild/moderate stroke and stochastically constrained following severe stroke," *Exp. Brain Res.*, vol. 143, no. 4, 2002, doi: 10.1007/s00221-002-1055-3.

[145] N. Shawen *et al.*, "Role of data measurement characteristics in the accurate detection of Parkinson's disease symptoms using wearable sensors," *J. NeuroEngineering Rehabil.*, vol. 17, no. 1, p. 52, Apr. 2020, doi: 10.1186/s12984-020-00684-4.

[146] J. Konrad, N. Marrus, and C. E. Lang, "A Feasibility Study of Bilateral Wrist Sensors for Measuring Motor Traits in Children With Autism," *Percept. Mot. Skills*, vol. 129, no. 6, pp. 1709–1735, Dec. 2022, doi: 10.1177/00315125221125275.

[147] M. K. O'Brien *et al.*, "Wearable Sensors Improve Prediction of Post-Stroke Walking Function Following Inpatient Rehabilitation," *IEEE J. Transl. Eng. Health Med.*, vol. 10, pp. 1–11, 2022, doi: 10.1109/JTEHM.2022.3208585.

[148] R. Sun, R. Song, and K. Tong, "Complexity Analysis of EMG Signals for Patients After Stroke During Robot-Aided Rehabilitation Training Using Fuzzy Approximate Entropy," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 22, no. 5, pp. 1013–1019, Sep. 2014, doi: 10.1109/TNSRE.2013.2290017.

[149] Y.-H. Wang, I.-Y. Chen, H. Chiueh, and S.-F. Liang, "A Low-Cost Implementation of Sample Entropy in Wearable Embedded Systems: An Example of Online Analysis for Sleep EEG," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–12, 2021, doi: 10.1109/TIM.2020.3047488.

[150] R. R. Bailey, R. L. Birkenmeier, and C. E. Lang, “Real-world affected upper limb activity in chronic stroke: an examination of potential modifying factors,” *Top. Stroke Rehabil.*, vol. 22, no. 1, pp. 26–33, Feb. 2015, doi: 10.1179/1074935714Z.0000000040.

[151] S. L. DeJong, S. Y. Schaefer, and C. E. Lang, “The need for speed: Better movement quality during faster task performance after stroke,” *Neurorehabil. Neural Repair*, vol. 26, no. 4, pp. 362–373, May 2012, doi: 10.1177/1545968311425926.

[152] C. J. Fisher, “Using an Accelerometer for Inclination Sensing,” *-1057 Appl. Note Analog Devices*, 2010.

[153] V. A. Swanson *et al.*, “Optimized Home Rehabilitation Technology Reduces Upper Extremity Impairment Compared to a Conventional Home Exercise Program: A Randomized, Controlled, Single-Blind Trial in Subacute Stroke,” *Neurorehabil. Neural Repair*, p. 15459683221146996, Jan. 2023, doi: 10.1177/15459683221146995.

[154] M. Yazdani, G. Gamble, G. Henderson, and R. Hecht-Nielsen, “A simple control policy for achieving minimum jerk trajectories,” *Neural Netw.*, vol. 27, pp. 74–80, Mar. 2012, doi: 10.1016/j.neunet.2011.11.005.