UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE FÍSICA

Wearable Technology and Machine Learning Algorithms to Monitor Upper-Limb Use in Brain Injury Survivors

Cristiana Isabel de Menezes Caixinha Ernesto

Mestrado Integrado em Engenharia Biomédica e Biofísica
Perfil em Engenharia Clínica e Instrumentação Médica

Dissertação orientada por:
Prof. Doutor Hugo Ferreira, Universidade de Lisboa
Prof. Doutor Paolo Bonato, Harvard Medical School

2022

Resumo

Há uma elevada incidência de alterações motoras nos membros superiores na sequência de lesões cerebrais adquiridas (LCA). Uma lesão cerebral pode acontecer devido à aplicação de uma força mecânica e/ou perturbação no fornecimento de sangue ao cérebro. Assim, estas lesões resultam de várias causas, sendo as duas causas mais comuns os traumatismos cranioencefálicos (TCE) e os acidentes vasculares cerebrais (AVC). As LCA afetam negativamente a capacidade do paciente em realizar atividades da vida diária (AVD), especialmente a manipulação de objetos, uma ação motora motivada pelo alcance da conclusão de uma tarefa. Esta investigação visa quantificar e monitorizar essas alterações motoras (i.e., resultados da reabilitação) em resposta a uma intervenção clínica com a duração de pelo menos seis semanas. Tradicionalmente, os resultados da reabilitação são relatados por um clínico, tornando a avaliação demorada e dependente do observador. Além disso, a avaliação não é feita regularmente, ou seja, na maior parte das vezes, as avaliações são feitas apenas em dois momentos: pré e pós-intervenção; é desejável fazerem-se avaliações ao longo de todo o plano de cuidados do paciente por forma a adaptar o plano de reabilitação de acordo com esses mesmos resultados. Assim, abordarei esta questão automatizando este processo através da derivação de estimativas de escalas clínicas que quantificam os resultados da reabilitação e são de elevado interesse clínico, utilizando dados de sensores inerciais e desenvolvendo algoritmos de Aprendizagem Automática.

O estudo foi conduzido numa amostra heterogénea de trinta e sete pacientes que sofreram lesões cerebrais e ficaram com sequelas motoras devido a essas mesmas lesões, nomeadamente hemiparesia de um dos membros superiores. Dezasseis destes pacientes eram sobreviventes de AVC, enquanto vinte e um eram sobreviventes de lesões cerebrais traumáticas. A literatura indicia que existe potencial para utilizar sensores em aquisições de dados como forma de obter uma avaliação da reabilitação mais precisa. Desse modo, será possível efetuar uma recolha de dados com estes sensores posicionados em diferentes configurações, o que, em conjunto, permitirá que os pacientes efetuem movimentos de forma menos constrangida aquando da performance das AVD. Além disso, os acelerómetros são os dispositivos mais utilizados para este fim, satisfazendo os requisitos do utilizador (por exemplo, precisão, conforto, baixo custo e facilidade de utilização), e tendo fatores relevantes (por exemplo, colocação, material, peso e dimensionamento) dado que é uma forma de tecnologia vestível. De salientar que estes dispositivos permitem aos investigadores e clínicos recolher dados no sentido de avaliar a recuperação motora em múltiplos pontos de tempo ao longo do programa de reabilitação, e, assim, o progresso da recuperação pode ser acompanhado mais de perto, sendo que decisões poderão ser efetuadas tendo em conta os resultados. Consequentemente, para permitir a avaliação dos efeitos das intervenções de reabilitação dos membros superiores, os pacientes foram instrumentados com um total de sete unidades de medição inercial (UMI) sem fios. Os pacientes foram guiados de forma a realizar uma série de oito tarefas padronizadas que foram especificamente escolhidas para reproduzir padrões motores encontrados em diversas atividades da vida quotidiana. As tarefas incluem a manipulação bimanual ou unimanual de objetos, ou ambos, e, o controlo motor fino ou grosso, ou ambos. Além do mencionado, foram desenvolvidos algoritmos de Aprendizagem Automática dedicados em derivar, em primeiro lugar, a Escala de Capacidade Funcional (FAS) relativa à qualidade de movimento do paciente durante o desempenho de atividades da vida diária e, em segundo lugar, a Avaliação Fugl-Meyer (FMA) especificamente para o membro superior de forma a medir a gravidade dos défices motores relativamente às lesões cerebrais adquiridas pelos pacientes. Para derivar a Avaliação Fugl-Meyer foram utilizados os conhecimentos aprendidos anteriormente com as estimativas da Escala de Capacidade Funcional, sendo que foi identificado que as duas escalas possuem uma correlação positiva.

Em termos gerais, as nossas conclusões revelam que para a Escala de Capacidade Funcional um modelo Hierárquico de Aprendizagem Automática com um Classificador Binário e Regressão no segundo nível é a melhor solução para estimar pontuações das escalas clínicas, com particular enfoque nos pacientes com maior capacidade motora. O nosso racional para este foco reside no facto de que clinicamente é de elevada importância efetuar este tipo de avaliação motora em pacientes que estão num nível relativamente elevado de função motora: estes pacientes têm a capacidade de gerar bons padrões fisiológicos de movimento e podem beneficiar mais amplamente do tratamento de reabilitação. O coeficiente de determinação, R^2, foi utilizado como métrica padrão para avaliar o modelo final devido à sua relevância no fornecimento de informações relativas ao ajuste do modelo em termo das observações das escalas clínicas estimadas em comparação com as observações rotuladas pelo clínico no laboratório. O modelo final da Escala de Capacidade Funcional alcançou um excelente desempenho com um coeficiente de determinação, R^2, de 0,91. As principais contribuições desta investigação relativa à Avaliação Fugl-Meyer sublinham que um subconjunto de tarefas emerge como ótimo para derivar a Avaliação Fugl-Meyer. Este subconjunto inclui tarefas que foram de particular dificuldade para a performance dos pacientes, tarefas de dificuldade média e também tarefas de fácil execução. Assim, em geral, este subconjunto compreende todos os níveis de aplicabilidade da qualidade motora, incluindo a manipulação bimanual e unimanual de objetos. Além disso, foi identificado que adicionar as pontuações da Escala de Capacidade Funcional (i.e., fornecendo conhecimentos relativos aos padrões de qualidade de movimento) ao modelo final da Avaliação Fugl-Meyer, tanto para tarefas individuais como para previsões combinadas, foi fundamental para melhorar o desempenho geral do modelo. Como resultado, o modelo final Avaliação Fugl-Meyer alcançou um bom desempenho com um R^2 de 0,79. Até à data, não existem estudos efetuados com este tipo de tarefas e esta metodologia hierárquica de forma a estimar a Escala de Capacidade Funcional proveniente de dados de sensores inerciais. Relativamente aos métodos utilizados para a automatização da Avaliação Fugl-Meyer, o nosso estudo é pioneiro. Portanto, a nossa análise é encorajadora, tendo demonstrado a viabilidade de incorporar dados de sensores adquiridos em tarefas explicitamente selecionadas e adaptadas para reproduzir padrões motores de atividades da vida diária e tirar partido do conhecimento derivado da estimativa da Escala de Capacidade Funcional para melhorar as previsões finais do modelo Avaliação Fugl-Meyer.

No seu conjunto, este estudo pode empoderar os profissionais de saúde na realização de reabilitação de precisão - uma intervenção ajustada de acordo com resultados de resposta específicos do paciente - maximizando assim os ganhos motores dos pacientes com lesões cerebrais adquiridas. Adicionalmente, como perspetiva futura, seria desejável avaliar não só a capacidade motora dos pacientes (avaliada na clínica), mas também o desempenho do paciente, isto é, o que os pacientes conseguem realmente fazer em ambiente doméstico e quando inseridos na comunidade. Isto é particularmente relevante uma vez que os pacientes com lesões cerebrais utilizam geralmente estratégias compensatórias no ambiente doméstico, ou seja, têm capacidade motora mas não a utilizam particularmente o membro hemiparético, o que faz com que seja de grande relevância fornecer um feedback constante, durante o plano de cuidados, de forma a que não aconteça uma perda de função do membro hemiparético. Assim, do mesmo modo, os resultados desta investigação abrem o caminho para a transição de resultados de reabilitação relatados pelos clínicos para estimativas automatizadas de resultados baseadas em sensores que, entre outros, permitem que os clínicos e cuidadores tenham instrumentos para que possam prestar serviços de cuidados à distância, nomeadamente em casa e na comunidade.

Palavras-chave: Lesão Cerebral Adquirida; Aprendizagem Automática; Reabilitação de Precisão; Tecnologia Vestível; Tecnologias Portáteis e Móveis para Reabilitação.

Abstract

There is a high incidence of upper-limb motor changes following an acquired brain injury (ABI). Those impairments negatively impact the patient's ability to perform activities of daily living (ADL), especially reach-to-grasp of objects, a goal-directed action. This research aims to assess and monitor those changes (i.e., rehabilitation outcomes) in response to clinical intervention. Traditionally the outcomes are clinician-reported, making the assessment time-consuming and observer-dependent. Also, the assessment is not done on a regular basis (i.e., most of the time, assessments happen only at two time points: pre and post-intervention), and it is desirable to assess over the cycle of care. Thus, this will be addressed by deriving estimates of clinical-based outcome measures using wearable sensor data and Machine Learning algorithms (ML).

The study was conducted on a heterogeneous sample of thirty-seven ABI individuals with upper-limb hemiparesis. Sixteen of these patients were stroke survivors, while twenty-one were traumatic brain injury survivors. Literature review shows that there is potential to use wearable technology to have a more precise assessment and collect data with different setups that will enable the performance of less constrained movements. Additionally, accelerometers are the most used devices for this purpose, meeting the user requirements (e.g., accuracy, comfort, low cost, and ease of use), and having relevant wearability factors (e.g., placement, material, weight, and sizing). Furthermore, these devices enable researchers and clinicians to collect data towards evaluating motor recovery at multiple time points, and thus, the recovery progress can be followed more closely. Consequently, to enable the assessment of the effects of the upper-limb rehabilitation interventions, subjects were instrumented with a total of seven wireless inertial measurement units (IMU). The subjects were guided to perform a series of eight standardized tasks specifically chosen to reproduce motor patterns of ADL. Dedicated ML algorithms were developed to derive firstly the Functional Ability Scale (FAS) concerning the patient's quality of movement during the performance of activities of daily living and secondly the Fugl-Meyer Assessment (FMA) for upper extremity (UE) assessment regarding the severity of motor impairments, where the knowledge learned from the preceding FAS estimations was employed.

Broadly translated, our findings reveal that for the FAS, a dedicated Hierarchical Random Forest-based model with a Binary Classifier and Regression on the second level is the best solution to derive scores with a particular focus on high motor functioning patients. The final FAS model achieved excellent performance with a coefficient of determination, R^2, of 0.91. The main contributions of this research pertaining to the FMA highlight that a subset of tasks emerges as the optimal to derive the FMA. Moreover, adding the FAS scores into the final FMA model for both single tasks and combined predictions was key to improving the performance by providing quality movement knowledge. As a result, the final FMA model achieved good performance with a R^2 of 0.79. Therefore, our analysis is encouraging, having demonstrated the feasibility of incorporating wearable sensor data acquired in tasks explicitly selected and adapted to reproduce ADL motor patterns and taking advantage of the knowledge derived from estimating the FAS to improve the final FMA model predictions.

Altogether, the methods herein presented have the potential to empower healthcare professionals to perform precision rehabilitation, an intervention adjusted according to patient-specific responsiveness, maximizing ABI patients' motor gains. Likewise, the findings of this analysis pave the way to transitioning from clinician-reported outcomes to sensor-based outcome estimates that, among others, can be assessed in home and community settings.

Keywords: Acquired Brain Injury; Machine Learning; Precision Rehabilitation; Wearable Technology; Portable and Mobile Technologies for Rehabilitation.

Acknowledgments

Firstly, I would like to express my most sincere and deepest gratitude towards my supervisors, Dr. Hugo Ferreira and Dr. Paolo Bonato. Their unconditional support, expertise, and guidance were crucial throughout this learning process. Thank you for all the precious advice and endless enthusiasm in sharing your extensive scientific knowledge in the field.

To Federico Parisi, my internal mentor, for in 2018, conducting my first interview to be part of the exceptional Motion Analysis Laboratory research team. I am grateful to you for believing in my capacities and encouraging me since day one and for the extraordinary constant constructive feedback and patience in the face of challenges during the development of this research.

I am likewise thankful to the entire members of the Motion Analysis Laboratory for being such good colleagues and a support system. Notably, I would like to especially thank Catherine, Gloria, Stefano, and Eric for the warm welcoming in sharing their workspace with me, for the inspiring clinical input, and for the friendship we grew over the past three years.

On a personal note, my decision to study Biomedical Engineering and Biophysics came about due to my fascination in the area of diagnosis and motor rehabilitation, motivated by a personal incident; being a stroke survivor at the age of twelve with idiopathic diagnosis. Hence, the Dissertation research topic speaks to my heart, and I will be forever grateful to my supervisors for the opportunity to contribute to improving the quality of life of people who sustained brain injuries and motor disability diagnoses. Furthermore, I deeply respect the acquired brain injury survivors who contributed to the MAL research. This experience is an incredibly important part of my education, and its teachings will stay with me throughout my life.

I gratefully acknowledge the scholarship grant received from the Luso-American Development Foundation (FLAD) to undertake my research in Boston. Particularly, I would like to thank Dr. Rita Faden for the sponsorship and Fátima and Ricardo for all the excellent advice and bureaucratic support.

Likewise, I wish to express my deepest gratitude towards the IBEB professors, especially Dr. Raquel and Dr. Nuno, who, along with my supervisors, showed unconditional support during the initial stages of this Dissertation when I had to overcome a cancer diagnosis; your kindness and encouragement will be forever in my heart.

I wish to extend my special thanks to the physicians Dr. Rita Santos and Dr. Helena Santos-Martins, for their dedication, thoughtfulness, and compassion.

Super smiling thanks to my best friends, Daniela, Diogo, Madalena, Mariana, Katrin, and Carolina, who continuously shared their uplifting words and unwavering faith with me; I most humbly appreciate you for being such inspiring, intelligent persons. I wish to thank my friends in Boston for being my family away from home, especially the Santos family, Liliana, Homa, and Zeynab. Thank you to São, Amélia and Marta, for being special rays of sunshine and for your wise words. I could not have undertaken this journey without you.

I extend my warmest thanks to my boyfriend, Jorge, for your endless companionship and inspiration in my life for over seven years; you are the one-of-a-kind special someone. To Jorge's parents, Helena and Jorge, and brothers João and Pedro, thank you for the constant steadfast encouragement and kindness.

A heartfelt appreciation to my family for being my safe haven and showing me unconditional love and encouragement throughout my life. To my parents, Ana Teresa and Marcos, for always believing in me and motivating me to fight for my personal and professional goals; you are an example of perseverance and fortitude. To my uncles, Valdo and João, for your love and care. To my grandparents, Isaura and João, for their lifelong wisdom, teachings, and kindness, which mean the world to me. I am eternally grateful to have you in my life.

Uma palavra de apreciação sincera à minha família por ser o meu porto seguro e por me mostrar amor e apoio incondicionais ao longo da minha vida. Aos meus pais, Ana Teresa e Marcos, por acreditarem sempre em mim e por me darem a coragem de lutar pelos meus objetivos pessoais e profissionais; vocês são um exemplo de perseverança e fortitude. Aos meus tios, Valdo e João, pelo vosso amor e carinho. Aos meus avós, Isaura e João, pela sua sabedoria, ensinamentos e bondade duradouras, que significam o mundo para mim. Estou eternamente grata por vos ter na minha vida.

Dedicated to my parents, Teresa and Marcos, and grandfather João.
Also dedicated to my late grandparents, Isaura, Elizabeth and Artur, and my paw brother Joy,
who flew away from this world too soon.

We keep on walking, kindly, on the shoulders of giants.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AAUT Actual Amount of Use Test

ABI Acquired Brain Injury

Acc Acceleration

ADASYN Adaptive Synthetic Sampling Approach

ADL Activities of Daily Living

AMAT Arm Motor Ability Test

AOU Amount of Use

ARAT Action Research Arm Test

BBT Box and Block Test

BI Barthel Index

CAHAI Chedoke Arm and Hand Activity Inventory

ELM Extreme Learning Machines

EMG Electromyography

FAS Functional Ability ScaleFugl-Meyer Assessment

FMA Fugl-Meyer Assessment

ICF International Classification of Functioning, Disability and Health

IMU Inertial Measurement Unit

IRB Institutional Review Board

JTHFT Jebson-Taylor Hand Function Test

KPF Knitted Piezoresistive Fabric sensor

LOSOCV Leave-One-Subject-Out Cross-Validation

MAL Motor Activity Log

MAS Modified Ashworth Scale

MEMS Micro-Electromechanical Systems

ML Machine Learning

MMSE Mini-Mental State Examination

MRMR Minimum Redundancy Maximum Relevance

NCA Neighborhood Component Analysis

NIH National Institutes of Health

OOB Out-of-Bag

PROs Patient-Reported Outcomes

QOU Quality of Use

RF Random Forest

RMA Rivermead Motor Assessment

RMS Root Mean Square

RMSE Root Mean Square Error

SD Standard Deviation

SE Standard Error

SIS Stroke Impact Scale

SL Stroke Lesion

SMOTE Synthetic Minority Oversampling Technique

SVR Support Vector Regression

t-SNE T-distributed Stochastic Neighbourhood Embedding

TBI Traumatic Brain Injury

TBL Traumatic Brain Lesion

UE Upper Extremity

UL Upper-Limb

VAS Visual Analogue Scale

WEKA Waikato Environment for Knowledge Analysis

WHO World Health Organization

WMFT Wolf Motor Function Test

Chapter 1

Introduction

This chapter aims to introduce the reader to the theoretical context of the work and its main objective. The background and motivation of the research are also emphasized. The complete structure of the dissertation is summarized and outlined in the final subsection.

1.1 Motivation and Background

1.1.1 Acquired Brain Injuries and Consequences

Acquired brain injury (ABI) is a term applied to describe a damage to the brain that is not hereditary, congenital, degenerative, or induced by birth trauma [18]. A brain injury can happen due to applying a mechanical force or/and disruption in blood supply to the brain. Thus, ABI results from various causes; the two most common causes are traumatic brain injuries (TBI) and stroke [19].

Globally, stroke is a leading cause of mortality and disability with an incidence of approximately 13.6 million persons in 2016 [20]. TBI is considered a global "silent epidemic" due to representing the greatest contributor to death and disability amongst any other traumatic events. The estimated global incidence of TBI, including all causes that led to TBI, is 69 million persons per year [21].

Furthermore, ABI is also one of the principal causes of death and long-term disability in the United States [13]. The American Heart Association estimates that there are 795 million new or recurrent cases of stroke each year (i.e., incidence) [14]. Studies on the epidemiology of stroke indicate that the chance of having a stroke increases with age, with the most significant risk factor being hypertension [18]. Due to the increase in life-expectancy that leads to an aging population, it is estimated that stroke prevalence will continue to increase [22]. Regarding TBI, the risk factors associated with a TBI injury vary between ages. Falls and sports accidents are the primary causal factor for persons under 20 years. Then, transport-associated accidents prevail amongst adults, and a high level of falls is noted for older adults [18]. Thus, the highest risk injury peaks are achieved at two different time stages; the first one is between 16 and 25, which will slowly decrease until late middle age and rise again to the second peak at around age 65 [18]. In the U.S. each year, approximately 1.4 million individuals sustain a TBI (i.e., incidence) [15]. Between 1970 and early 1980s, researchers estimated that more than 500 thousand persons were hospitalized or died due to TBI each year [23]. Concerning TBI, there is difficulty noted in acquiring accurate measures of prevalence and incidence due to disparities in definitions and assessments of collecting data [15]. A summary of these statistics and further information on the burden costs associated with ABI can be seen in Table 1.1.

Table 1.1: Incidence, prevalence, costs, and impact on disability of ABI in the US. Adapted from [13], [14], [15]

Condition	Prevalence	Incidence	Annual Total Cost	Activity Limitation
Stroke	6.8 million adults aged ≥20y; 2.8% of adult population	795 thousand persons/year	$36.6 billion (in 2013 dollars)	Among stroke survivors aged ≥65y, 26% were dependent in activities of daily living, 50% had hemiparesis, 30% unable to walk without assistance, 19% had aphasia, 26% were in a nursing home 6 months post-stroke
TBI	3.32 million with long- term disability; 1.1% of total population in 2005	1.4 million persons/year	$78.1 billion (in 2013 dollars)	43% of persons discharged after acute TBI hospitalizations develop long-term disability

Many unknowns revolve around the nervous system function. Following a brain injury, clinicians cannot readily predict outcomes for each patient [24]. ABIs can bring permanent disabilities or functional deficits resulting in long-term care needs for patients. Amongst the many sequelae of ABIs are cognitive, behavioral, psychological, and motor impairments. In this project, the focus will be on the upper-limb (UL) motor impairments associated with ABI.

The use of the upper-limb is needed to perform most of the activities of daily living (ADL). Particularly, reach-to-grasp of objects, a goal-directed action, is often mentioned as a key function of a healthy upper-limb [25]. In the long-term, it is suggested that persistent motor impairments (i.e., permanent or temporary problems in the function or structure) of the upper-limb amongst stroke and TBI survivors have a prevalence between 29% to 88%, following neurological injury [26], [27]. Concerning TBI, there is a range of combined or separate motor impairments of the upper-limb that depend on the site and damage extension following the trauma [18]. Upper-limb hemiparesis (i.e., weakness or paralysis of muscles on one side of the body) and functional effects on the paretic (i.e., compromised) arm and hand are frequent stroke sequelae; nevertheless, hemiparesis can also be present in TBI patients.

As depicted in Figure 1.1, hemiparesis constitutes daily coordination limitations to object reach (arm and hand moving towards an object) and manipulation (hand opens and closes on the object) [25], [28], [29]. Consequently, upper-limb impairments can contribute to a loss of independence in daily living activities, reducing these patients' quality of life. Even though motor nonuse may initially occur due to weakness, paralysis, or sensory loss, in the long term, the person can become acquainted with not incorporating the paretic limb in functional activities, even though the individual can move it, a phenomenon known as learned nonuse [1]. Ultimately, this results in a motor form of negative compensatory reinforcement strategy. Hence, there is a necessity to tackle the high prevalence of long-term upper-limb disabilities resulting from ABI. The identification of rehabilitation processes that aim to restore the upper-limb function is crucial to prevent learned nonuse, accelerate the recovery of these upper-limb impairments, and compensate for the disabilities. With rehabilitation, the ultimate goal is for the patient to be able to perform key activities in their daily life with their affected arm.

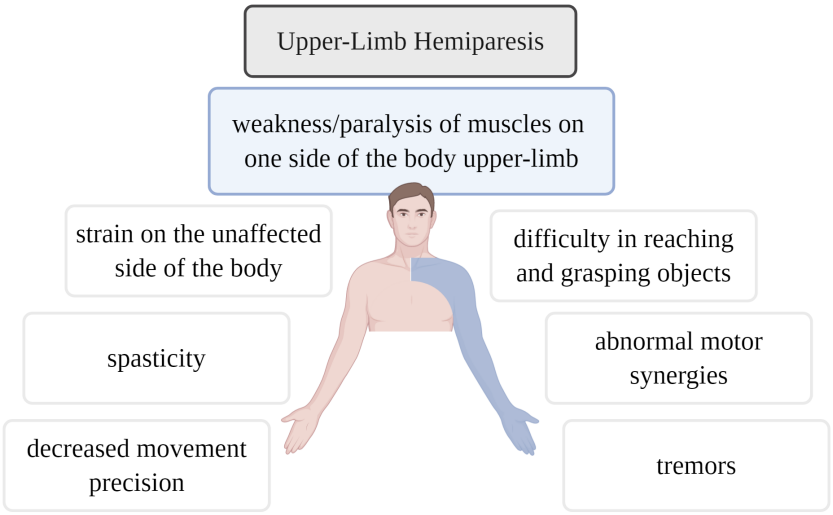


Figure 1.1: Hemiparesis definition and respective side effects [1]. The hemiparetic side of the body is represented in the shaded blue area of the human anatomical representation.

Patient Assessment According to the International Classification of Functioning, Disability and Health

A comprehensive assessment of brain injury survivors is crucial to delivering a suitable rehabilitation plan that promotes motor gains.

Rehabilitation is organized around the International Classification of Functioning, Disability and Health, a clinical tool commonly known as ICF, that was developed by the World Health Organization (WHO) [2]. This framework enables clinicians to consider the patient as a whole across the different interconnected domains. Thus, the utilization of ICF is increasingly promoted amongst researchers and clinicians. Overall, this enables an international standardized acquisition of levels of functioning and disability, variables whose data are vital and is currently lacking in everyday rehabilitation practice [30]. This classification system is mentioned as having the potential to drive a paradigm shift in health and disability [30].

The main dimensions of the ICF model (Figure 1.2) are body functions and structures, activity, and participation. Additionally, external and internal factors are considered (i.e., respectively, environmental and personal factors). Figure 1.2 depicts the classification system applied to ABI.

ICF system can be utilized to guide the development, use, and interpretation of patient-reported outcomes (PROs) and performance-based outcome measures (i.e., clinician-rated outcome measures) [31]. PROs are outcomes reported by the patient and usually aim to know more about the overall functioning and sense of wellbeing. Performance-based outcome measures are reported by healthcare professionals based on the observation and grading of a patient's health condition. These outcomes are highly dependent on clinical judgment and interpretation of a patient's detectable behaviors, signs, and physical manifestations. The two approaches can be used to acquire data from the ICF system's main dimensions. As depicted in Figure 1.2, these dimensions are interconnected; therefore, to evaluate treatment effects and manage clinical interventions, everything must work collectively.

The dimension of body functions and structure (e.g., upper-limbs) aims to identify the patient's physiology and psychology following the ABI. To delineate ABI patients' motor rehabilitation plan, physical assessments, and PROs allow clinicians to have an idea of the patient's impairments, difficulties (e.g., at home), as well as their abilities and aspirations. The definition of the patient's main physical

problems is based on assessing movement dysfunctions'. Hence, this field takes into consideration that deviations from a natural function and structure should be classified as impairments. This approach is challenging to replicate amongst therapists because identifying these problems depends on the therapist's background, clinical judgment, and time available [4]. This method can potentially result in inaccuracy when measuring the improvement or decline in motor gains.

Activities (i.e., execution or performance of an action or task) include representational assessments that aim to indicate the patient's function level [16]. Thus, regarding ABI, it is vital to perform motor functioning assessments. This type of assessment can be performed in different approaches. Regularly, activities such as motor task performance are used as outcome measures. Nevertheless, it is important to note that task completion does not always reflect recovery (e.g., patients can adopt motor patterns that compensate for lost limb function) [4]. Therefore, it isn't easy to differentiate between motor restitution and compensation strategies with these outcome measures.

Participation is regarded as how the patient deals with and is involved in daily life situations after the ABI. Personal and environmental factors can influence this participation. Thus, it is important to incorporate disability as the term that includes persons with impairments, activity limitations, and participation restrictions [18].

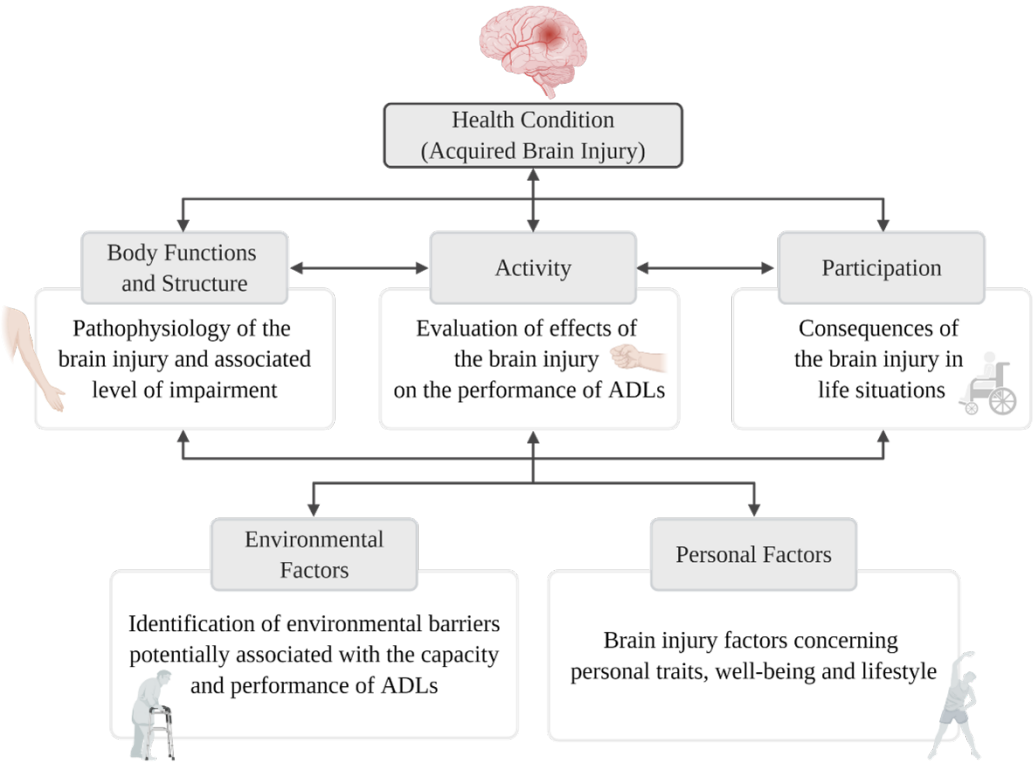


Figure 1.2: International Classification of Functioning, Disability and Health adapted to highlight the principal components for ABI health condition [2].

Rehabilitation Process

Following an event such as a TBI or stroke, assessments are necessary tools to plan and then adjust the rehabilitation intervention (Figure 1.3). Nevertheless, assessments and interventions are interconnected and practiced over the care cycle, usually beginning in the intensive care unit (i.e., before rehabilitation)

and continuing with outpatient visits. To do the assessment of the patient on a regular basis is essential to evaluate the rehabilitation intervention [32]. However, traditionally these outcome measures are essentially collected at two time points: baseline and discharge. This is limitative because compromises the maximization of motor gains, making clinical intervention adjustment challenging to perform. Thus, as depicted in Figure 1.3, these assessments should ideally be done longitudinally (i.e., over the cycle of care) so that it is possible to monitor the clinical outcomes consistently in an unobtrusive way.

There is potential for precision-rehabilitation that can facilitate the estimation of clinical outcome measures that might indicate a progression or deterioration of the motor capabilities. In this way, clinicians can identify and target patient-specific motor deficits during rehabilitation planning. Hence, the rehabilitation plan and the techniques to achieve better clinical outcomes could be chosen accordingly. Therefore, these measures need to be easy to gather and easy to interpret.

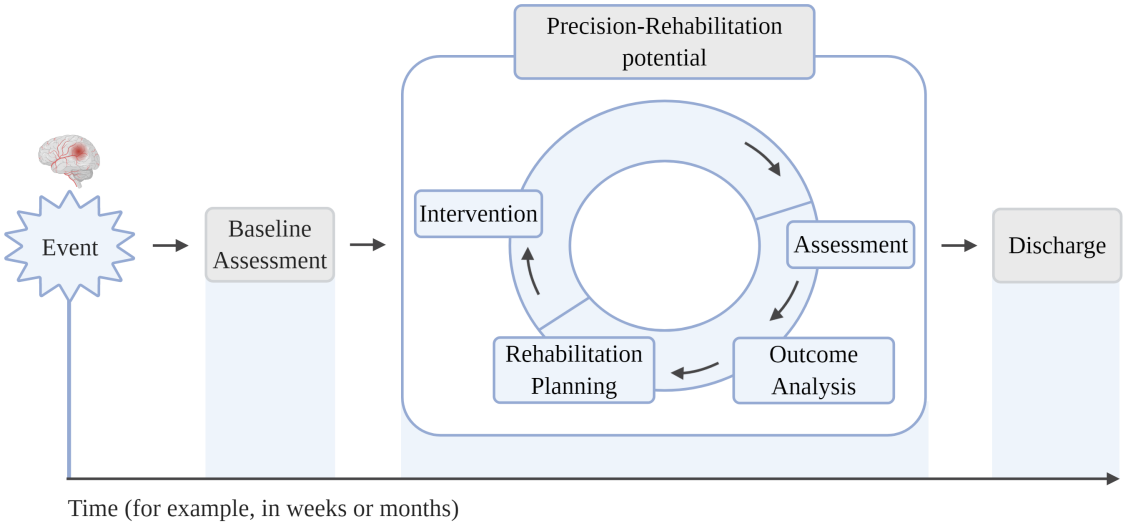


Figure 1.3: ABI motor rehabilitation phases from a patient point of view.

The use of outcomes to track motor capabilities has the potential to include a systematic recording and review of ABI patient's recovery status. Also, there is a need to assess the patient's function at home and find solutions to assess outside the clinic. Therefore, it is desirable to develop novel technologies of motor outcome measure systems with these features, guaranteeing reliability, sensitivity, and validation.

Evaluating Motor Recovery at Multiple Time Points

Wearable sensors and powerful Machine Learning algorithms are technologies that have the potential to enable researchers and clinicians to collect data towards evaluating motor recovery at multiple time points (i.e., also during the course of therapy) [33], [34]. Thus, the recovery progress can be followed more closely as opposed to solely before/after an intervention. Nevertheless, there are other long-term benefits that result from the use of these tools, such as allowing clinicians to spend less time doing assessments and offering technology that will enable data to be recorded in home and community settings during the performance of ADL. The benefits of tracking patients' recovery at home settings are that often, what patients are capable of doing in the clinic does not reflect how patients perform while at home.

Among wearables devices, one of the most commonly used categories is the one focusing on sensors that quantify patient movements. These sensors provide the capability to characterize motor biomarkers (e.g., movement time), which can be used to identify how well the body responds to an intervention,

distinguish mechanisms of restitution and compensation recovery, perform telerehabilitation and remote monitoring [35].

There are distinct ways of collecting data at different time points using wearable devices that will be in distinct ways relevant to assessing appropriate motor outcomes [24]. Therefore, to have more data points, monitoring patients over extended periods of time during the performance of ADL or home-based exercises with the use of wearable devices can be implemented. Furthermore, a broad spectrum of sensor placements can be chosen depending on the information desirable to gather. The patient's impairments, functional limitations, and quality of use are sections of interest regarding motor outcome analysis.

Nowadays, Machine Learning algorithms are widely used in medical applications and research projects such as medical diagnosis and treatment [36], [34]. Particularly, algorithms that automate decision-making processes by generalizing from known examples can be used as a base to extract clinically relevant motor outcomes from data collected using wearable sensors [37], [7]. These algorithms are known as supervised learning algorithms because they have the capability to learn from input/output pairs [38]. Hence, these algorithms have the potential to combine information from multiple movement parameters and estimate clinically relevant information .

1.2 Aim of the Dissertation

There is a high incidence of upper-limb motor changes following ABI (Figure 1.4). Those impairments have a negative impact on the patient's ability to perform activities of daily living.

The aim of this research is to automate the assessment and monitoring of ABI patients' upper-limb motor changes (via rehabilitation outcomes) in response to a clinical intervention. This can be addressed in a relatively unobtrusive way by deriving estimates of clinical outcome measures with the use of wearable sensor data and the development of Machine Learning algorithms. Particularly, for this purpose, clinical scales (i.e., outcome measures) pertaining to the patient's quality of movement (FAS) and severity of motor impairments (FMA) were assessed.

The approach conducted in this Dissertation is innovative since it focuses on incorporating wearable sensor data acquired in tasks explicitly selected and adapted to reproduce ADL motor patterns with novel ML methodologies applied in a heterogeneous ABI cohort. Furthermore, the knowledge derived from estimating the FAS is leveraged to further improve the final FMA model predictions.

This approach can lead to several advantages, including providing healthcare professionals with the tools to perform precision rehabilitation (i.e., intervention adjusted according to each individual's outcomes) and hence potentially maximizing patient motor gains.

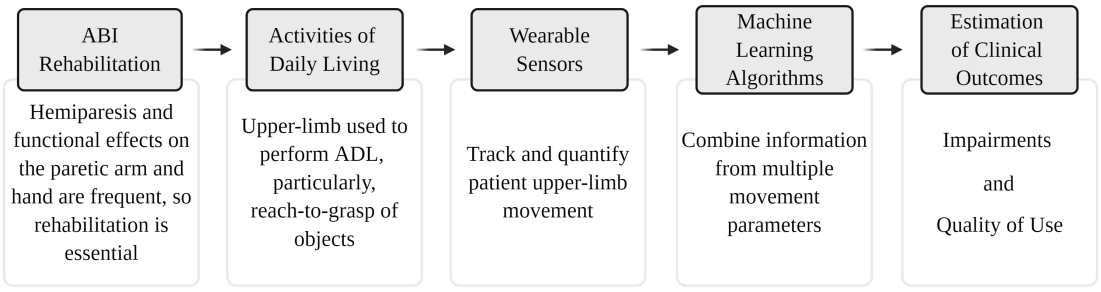


Figure 1.4: Summary of the main components of the research.

1.3 Dissertation Outline

This dissertation is organized into six chapters that convey the work conducted at the Motion Analysis Laboratory, Department of Physical Medicine and Rehabilitation of Harvard Medical School at Spaulding Rehabilitation Hospital in Boston. A brief overview of the contents included in each chapter is described as follows:

- Chapter 1: **Introduction** in which the underlying theoretical context of the work is presented. The motivation and background which drive the research are emphasized. The dissertation objectives and final aim are described. The complete structure of the dissertation is summarized and outlined.

- Chapter 2: **State of the Art** leverages the literature review in order to address the theoretical foundation and relevance of the topic. In this chapter, three main topics are explored and revisited: clinical scales for upper-limb assessment and their limitations, upper-limb movement kinematics, and automation of clinical score estimates. Eventually, in the final section, the relevant considerations of the review are scrutinized.

- Chapter 3: **Materials and Methods** introduces firstly the procedures involved in the laboratory data collection; here, the clinical information concerning the participant's demographics and inclusion criteria are described to provide a comprehensive view of the patient's health. Furthermore, are explained the study design and the technical specifications of the IMU sensor and its operation. Finally, the FAS and FMA clinical scales used to assess the patient's motor capabilities are presented. The second part of this chapter is dedicated to leveraging the *rationale* behind the Machine Learning algorithm development, where are explained the specific techniques and models used to analyze and derive the clinical scores, ranging from the signal processing techniques to the feature extraction and selection and, ultimately, the ensembled ML algorithm.

- Chapter 4: **Results** presents the results accomplished in the analysis of a dataset consisting of a total of thirty-seven acquired brain injury patients. The study investigated the two timepoints, Baseline, and Post-Treatment data assessments. The main findings regarding the participant's clinical data are described. Furthermore, the pre-processing methodology outcomes are underlined. Ultimately, the FAS and FMA resultant analysis of the estimated clinical scores focused on the upper-limb assessment are presented in individual sections. Finally, the models evaluated to predict the single and final clinical scores are examined.

- Chapter 5: **Discussion** highlights a thorough critical analysis concerning the main findings of the new automation techniques developed to assess ABI clinically relevant motor outcomes, namely FAS and FMA scores. Ultimately, this section emphasizes the feasibility of this analysis in providing the healthcare professional with novel tools to assess upper-limb rehabilitation outcomes more consistently over the care program and tailor the patient-specific plans accordingly. The final chapter section provides an overview of the main identified study limitations.

- Chapter 6: **Conclusions and Future Work** aims to summarize the main conclusions of the Master's Dissertation and provide a personal reflection on the objectives that were achieved during the course of the study. It also explores the various opportunities that future research can contribute to in the field of upper-limb-focused motor rehabilitation of Acquired Brain Injury patients.

Chapter 2

State of the Art

This chapter explores the various aspects of the previous work performed in this research field and develops a comprehensive understanding of its theoretical foundations and relevance. Three main topics are revisited: clinical scales for upper-limb assessment and their limitations, upper-limb movement kinematics, and automation of clinical score estimates. In addition, in the final section, the relevant considerations of the presented review are scrutinized.

2.1 Clinical Scales For Upper-Limb Assessment and Their Limitations

Several clinical scales have been proposed with the aim of assessing a patient's responsiveness to a rehabilitation intervention. The ICF framework enables the classification of these scales into three categories (i.e., body structure and functioning, activity, and participation). The most currently used scales to assess the upper-limb in rehabilitation are presented in Table 2.1.

Table 2.1: Summary of the most common clinical scales used in clinical settings to classify upper-limb outcome measures based on the ICF components. The categories: Body structure and Functioning, Activity, Participation represent the classification of the upper-limb outcome measure based on the component of the ICF.

Clinical Scales	Body structure and Functioning	Activity	Participation
Fugl-Meyer Assessment (FMA)	✓		
Arm Motor Ability Test (AMAT)	✓		
Visual Analogue Scale (VAS)	✓		
Modified Ashworth Scale (MAS)	✓		
Wolf Motor Function Test (WMFT)		✓	
Motor Activity Log (MAL)		✓	
Action Research Arm Test (ARAT)		✓	
Rivermead Motor Assessment (RMA)		✓	
Functional Ability Scale (FAS)	✓	✓	
Box and Block Test (BBT)		✓	
Barthel Index (BI)		✓	
Chedoke Arm and Hand Activity Inventory (CAHAI)		✓	
Stroke Impact Scale (SIS)			✓
Jebson-Taylor Hand Function Test (JTHFT)	✓	✓	

Additionally, the scales mentioned comprise differences, namely in the number of items, tasks included (unimanual/bimanual), activities of daily living included (basic = necessary to daily self-care, including personal hygiene, dressing, feeding, toileting, functional transfers; mobility/extended = home maintenance and required for independent living.), upper-limb use (AOU = amount of use; QOU = quality of use). Besides this, it is often valuable to consider the scale's validity, reliability, and responsiveness for the benchmark analysis of the clinical scales relevant to assess [7].

The Fugl-Meyer assessment (FMA) scale [39] is one of the most widely used to evaluate the severity of motor impairments. Specifically, the FMA-UE consists of assessing Upper-Extremity (UE) motor function. This scale includes a score between 0 and 2 that is assigned to each exercise (See Appendix C). However, it is mentioned in the literature that some scales as the FMA, are not able to completely capture the spectrum of the patient's motor function due to low sensitivity and ceiling effects [4].

The WMFT is an extensively used score that enables the quantification of upper-limb motor ability during functional tasks (generally includes 17 items) [40]. The WMFT is timed and also evaluated according to the Functional Ability Scale (FAS) pertaining to the performance of the UE [40]. The FAS consists of a clinician-reported score between 0 and 5 given to each exercise (visually rated) for a total score of 75 points, aiming to assess activity, and body structure and functioning (See Appendix B).

Another clinical scale often used is the Motor Activity Log (MAL). The MAL is an assessment that provides a patient-reported outcome concerning the quality and amount of movement of the more-impaired arm (scored in a 6-point ordinal scale) during 30 daily life functional tasks [41].

Nevertheless, the practical use of these scales is very time-consuming for the healthcare professional, and there are limitations noted in the objectivity and test responsiveness to subtle changes. Additionally, there is an increased yearning to have a rehabilitation precision medicine approach that enables the clinician to analyze the patient's particular impairments, functional limitations, and the quality of the rehabilitation measured with instrumental devices.

2.2 Upper-limb Movement Kinematics

To accurately assess clinical rehabilitation treatments is necessary to quantify the rehabilitation progress. Kinematics, in this context, is the study of the motion of the upper-limb. With kinematics, there can be provided precise measures of the upper-limb three-dimensional movements [35]. There are different ways of collecting upper-extremity movement data in three-dimensions that will be in different ways relevant to assessing appropriate motor outcomes from ABI patients [24]. Two fundamental approaches that can be used to this aim: reconstructing movement trajectories and extracting the movement patterns (i.e., features) characteristics directly from the sensor data. There is potential to identify motor biomarkers and differentiate between restitution of motor function and compensation motor recovery mechanisms with the use of these measurements [35].

2.2.1 Optometric Systems

The reconstruction of movement trajectories can be done using varied techniques; one of the most used are the optical motion capture systems [5].

A study by Hingtgen *et al.* used an upper-extremity kinematic model for the evaluation of the hemiparetic limb of stroke patients with spasticity during the performance of reaching tasks. The data was collected using passive optical motion capture technology [42]. This system used retroreflective markers placed in the segments of the affected and unaffected upper-limb. These markers could be tracked by in-

frared cameras, whose, with the proposed kinematic model of motion analysis, will enable the derivation of kinematic variables outcome measures such as movement time, range of motion, and peak angular velocity. Overall, the results of this study were indicative that the model could accurately quantify upper-limb arm motion.

Numerous studies have examined stroke patients during the performance of reaching tasks in two-dimensions and tried to correlate that with clinical outcomes; however, few studies have leveraged the correlation between three-dimensional reach-to-target kinematics and clinical outcomes [43],[44],[45].

A recent study by Adans-Dester *et al.* investigated the relationship between kinematic parameters of 3D reach-to-target movements and upper-limb clinical outcome measures in response to robot-assisted therapy [3]. The study included ten chronic stroke survivors. A clinician assessed the patient at baseline and the end of the intervention using the following clinical scales: Fugl-Meyer Upper-Extremity Assessment, Wolf Motor Function Test, and Functional Ability Scale. To assess patients based on the kinematic parameters derived from a biomechanical model was necessary to place markers on the patient's upper-limb, and a 10-camera motion capture system was used to track the movement. The patients were instructed to perform thirty-six reach-to-target movements with the paretic limb at a self-selected speed. The target was a panel with specific numbers (Figure 2.1 a.). Finally, a biomechanical model was then constructed (Figure 2.1 b.) in order to derive kinematic parameters of interest, namely, for example, the movement time (indicative of efficiency), trajectory directness (indicative of accuracy), peak velocity (indicative of speed), and time to peak velocity (indicative of planning of the movement).

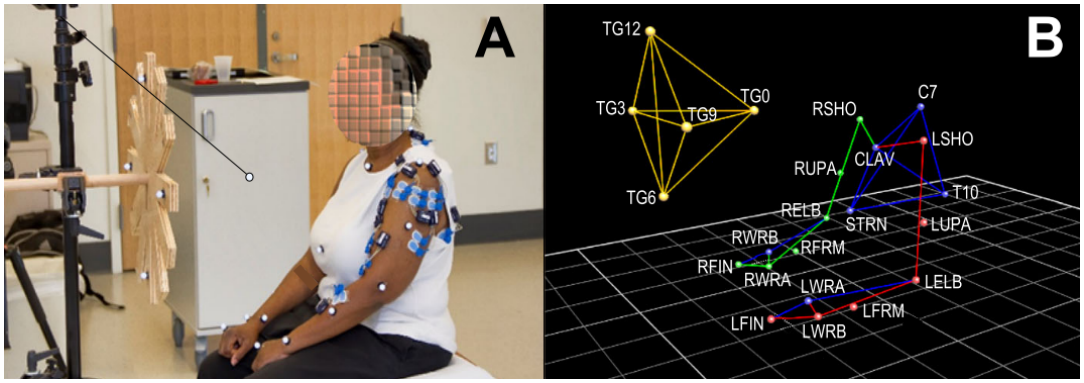


Figure 2.1: Experimental Set-up. a. Subject set-up: Twenty reflective markers were placed on body landmarks. b. Biomechanical model of subject in 1a and the target panel: Biomechanical model (Plug-in-gait) applied to reconstruct UE segments and derive kinematic parameters. Reproduced from [3].

The results of the study indicated a moderate-to-strong correlation between clinician-reported outcome measures and specific kinematic parameters in response to the intervention. Furthermore, the correlation between changes in clinical outcomes and kinematic parameters in response to the intervention differed widely across participants and was not strong across kinematic parameters. This suggests no consistent alteration in upper-limb motor strategies across participants, revealing a need to investigate further, at an individual level, about the response to interventions. Thus, the next step would be to identify similar motor patterns from different individuals and create clusters accordingly, enabling an opportunity to have cluster-specific kinematic parameters representative of clinical outcomes. Still, it was noted that it would be interesting to use clinical scales that include standard sets of tasks and, in this way, enable kinematic data to be aggregated accordingly. Furthermore, it was remarked that wearable sensors could enable upper-limb measurements during daily living activities in other settings providing a better understanding of the clinical outcome measures.

Marker-based optical measurement systems have limitations, such as using a rigid-body model for approximation of the UE, the marker-set movement with respect to the bones, and modifications of the marker-set placement positions when performing follow-up assessments, the high costs associated with these systems. Also, this technique is very time-consuming, and is necessary training to use these systems.

2.2.2 Wearable Sensors

Extraction of movement pattern characteristics directly from the sensor data is a technique that has been increasingly used in physical medicine and rehabilitation due to its practicality. Over the past years, studies that assess ABI upper-limb movement impairments have gained researchers' and clinicians' interest and have been progressively developed [5]. There is a wide range of wearable-sensing devices that can be used to acquire upper-limb movement data from ABI patients, namely: inertial measurement unit (IMU), accelerometer (Acc), flexible sensors, surface electromyography (EMG), magnetometer, encoders, e-textiles (e.g., knitted piezoresistive fabric sensor (KPF), stretch sensing fabric) are some wearable sensors frequently used to this purpose (Figure 2.2) [5].

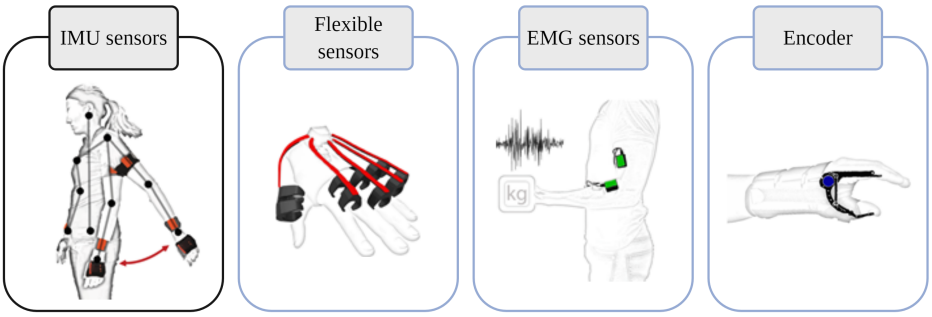


Figure 2.2: Examples of wearable-sensing devices that can be used to acquire upper-limb movement data from ABI patients. IMU sensors (orange) are the most used to track UL movements in ABI patients. The sensor set comprises: hands, forearms, and upper-arms; Flexible sensors (red) placed along the fingers. Piezo-resistive changes are detected in response to flexion; EMG sensors (green) sensors capture electrical potential difference as the depolarization wave travels along the muscle cells' membranes. The resulting signal is filtered and amplified for further processing. The sensor set is: over biceps and flexor digitorum superficialis muscles, involved in the elbow and wrist flexion, respectively; Encoder (blue) to track angular displacement of fingers supported by the orthosis where the encoder is placed (aligned with the rotation axis of the index finger). Icons available at [4].

A review by Wang *et al.* [5] leveraged studies that used interactive wearable systems for upper-limb rehabilitation. The primary aim of the research was to make an inventory of interactive wearable systems for movement and posture monitoring during upper-limb rehabilitation, as well as to classify the studies by the sensing technology, system measurements, and feedback conditions. Another goal was to leverage the currently available wearable systems and identify supporting information on their effectiveness and their wearability. The authors concluded that amongst the forty-five papers evaluated, accelerometers and inertial measurement units were the most frequently used technologies in the studies (84% of the papers). Note that some authors of the research papers analyzed consider accelerometers in the category of the IMUs, so they don't specify the category of the IMUs (Figure 2.3). Furthermore, as depicted, it is possible to interpret that IMUs technology generally fits all the boxes regarding measurement and feedback categories. Among all sensing technologies, the studies' feedback differed with a demand for visual and multi-modal feedback. The visual display is the most common form of feedback, enabling users to learn a motor task with an intervention that includes an instruction to be achieved or allowing the patient to compare to the correct/desired movement.

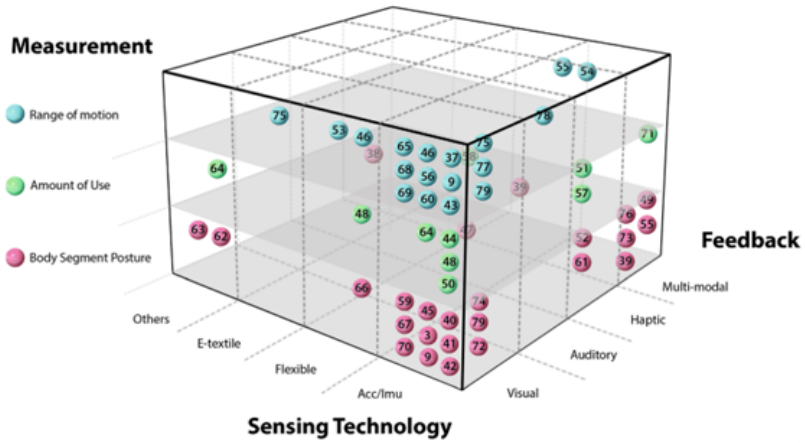


Figure 2.3: Taxonomy of interactive wearable systems regarding sensing technology, system measurement and feedback modalities. Reproduced from [5].

The focus of these studies was predominantly on stroke rehabilitation with the aim to monitor and provision of feedback on posture and upper-limb movement. This was possible by measuring upper-limb segment joint kinematics and/or trunk posture with different sensor configurations. Thus, the measurement classification used was the range of motion, meaning the movement distance around joint or body part, amount of use, which refers to the activity amount of body segment, and body segment posture to target spatial location. The final goals included overcoming learned non-use, improving the patient's limb range of motion, and performing ADL skills. Overall, for these particular goals, even though there are other options available, IMUs are considered the best sensing technology option due to having characteristics that are of interest to our analysis of these patients.

IMUs are essentially force sensors that consist of tri-axis gyroscopes and accelerometers. The gyroscopes enable the detection of the angular motion around one or several axes (turning rate), and the accelerometers provide acceleration readings along one or several axes [46].

The principle of operation of an accelerometer relies on measuring electrically physical changes in mass displacement, which can be explained by Newton's Second Law. Thus, a mass attached to a suspension structure with a reference frame will be subject to an inertial force due to the acceleration of gravity that will cause the mass to deflect. Different types of accelerometers have been developed to measure these changes, with piezoresistive, piezoelectric, and differential capacitive accelerometers being the commonly used [46]. Depending on the application of these sensors, different sample rates can be chosen such that aliasing (i.e., Nyquist rate, two times the frequency of the studied phenomenon) is prevented.

A gyroscope enables the measurement of the angular velocity rate using the Coriolis effect. There is a mass that moves in a particular direction with a particular velocity; when an external angular rate is applied, a force will occur, which will cause perpendicular displacement of the mass. Therefore, the displacement will cause a change in capacitance, which will correspond to a particular angular rate [46]. Thus, the orientation of the device could be obtained by integrating the angular velocity from the gyroscope. However, this leads to integration drift of the estimated orientation over time (e.g., due to gyroscope bias). In order to achieve a better estimation of the device orientation, a magnetometer (i.e., magnetic sensor) can sometimes be found in addition to an IMU unit [4]. The benefit of including a magnetometer, which is not an inertial sensor, is to have a reference on the orientation of the attached mass relative to the Earth's magnetic field.

A reliable estimation of the device orientation can be obtained by fusing the data recorded by the gyroscope, the accelerometer, and the magnetometer, as shown in Figure 2.4. Once the orientation of the device (and the associated body segment) is computed, this information can be used to remove the gravity component from the signal and obtain linear acceleration. Finally, by using double integration and specific filtering techniques, the position of the device in the 3D space could be derived [6].

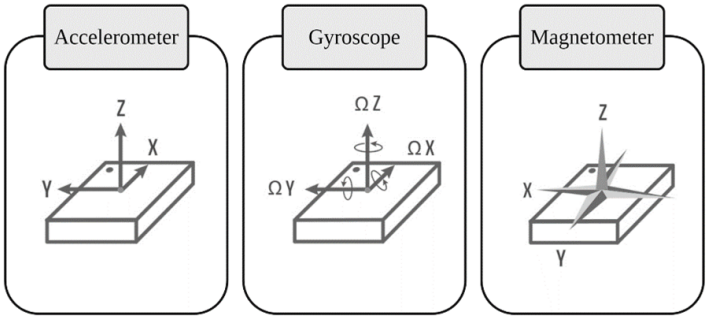


Figure 2.4: Representation of the IMUs components (Accelerometer and Gyroscope) and a additional component (Magnetometer) used to extract upper-limb kinematics data. Adapted from [6].

In most studies that track UL movement from ABI patients, it is often only necessary to use accelerometers because the goal is usually not to reconstruct exactly the limbs' tridimensional position to understand how the person is moving. Thus, in some studies, only accelerometers are used to capture the subjects' overall motor behavior (no orientation). This enables the ability to record data for a longer period of time and reduce data processing complexity. Therefore, accelerometers provide acceleration readings, but they usually include gravity, so it is not linear acceleration. However, accelerometers enable the possibility to have an approximation of the linear acceleration by using filters (for example, a high pass filter) to remove gravity, rather than using orientation estimation techniques to do so.

A study by Moreira *et al.* used eleven IMUs attached to a glove to perform hand and finger tracking with the aim to rehabilitate hand function of patients that had neurological disorders (such as ABI) related impairments [47]. This approach captures the hand movement with high sensor accuracy and stability over time. However, the conclusions on finger flexion trajectories evaluation in four different tasks were still within satisfying limits, indicating that additional IMU units might be needed to improve motion capture. Overall, this research provided a reliable alternative to traditional clinical hand function assessment.

IMU or accelerometer wearable-sensing systems enable the measurement of kinematic parameters of interest regarding ABI, such as orientation, position, velocity, and complex body posture and joint range of motion [5]. Advances in micro-electromechanical systems (MEMS) have facilitated the development of miniaturized accelerometer, magnetometer, and gyroscope designs. When utilizing IMU sensors for finger movement monitoring, the number of sensors needed to be used varies amongst studies, as well as different sensor placement approaches can be employed. Figure 2.5 illustrates different sensor placements that were used in the studies analyzed by Wang *et al.*.

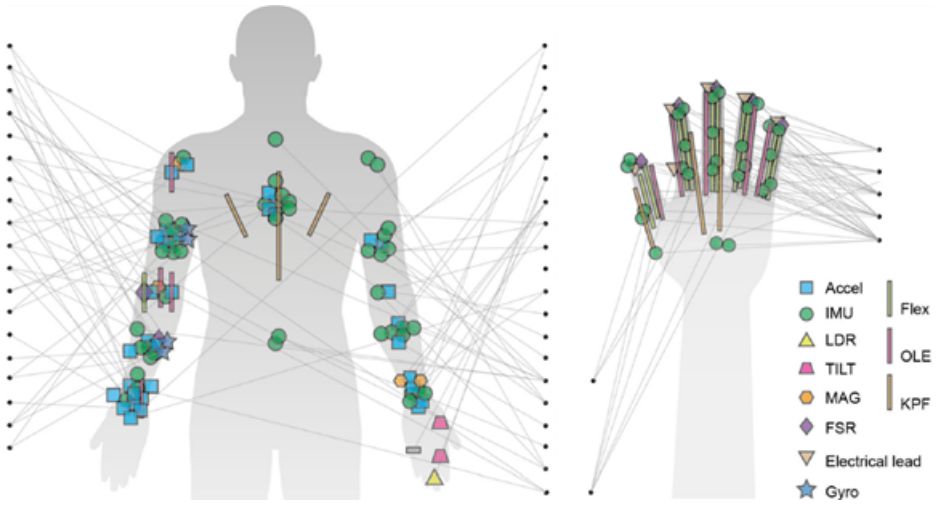


Figure 2.5: Sensor placements infographic. Each dot represents a study, the symbols depict different sensors used. Adapted from [5].

Additionally to the different approaches used in terms of sensor placements, several studies also use distinct task-based assessments such as standardized tasks vs. ADLs or unimanual vs. bimanual tasks. Table 2.2 describes an overview of the most common clinical scales that are used together with IMU technology to study ABI patients' rehabilitation progress.

Table 2.2: Summary of measurement instruments and studies involving the use of IMU technologies to study ABI patients. The studies were classified depending on the tasks being performed (unimanual and/or bimanual), the activities of daily living being assessed (basic ADL - activities of daily living necessary to daily self-care, including personal hygiene, dressing, feeding, toileting, functional transfers, and mobility; extended ADL - activities of daily living, beyond basic ADL, related to home maintenance and required for independent living). The validity, reliability, and responsiveness of the clinical scales used were evaluated based on the literature review. Adpated from [16], [17].

Measurement Instrument	Unimanual Tasks	Bimanual Tasks	Basic ADL	Extended ADL	Validity	Reliab.	Responsiven.	IMU
FAS	✓	✓	✓	✓	[48]	[48]	[48]	✓
FMA	✓	✓	✓	✓	[49]	[49]	[49]	✓
WMFT	✓			✓	[50]	[50]	[50]	✓
MAL	✓	✓	✓	✓	[51]	[51]	[51]	✓

2.3 Automation of Clinical Score Estimates

In the next subsections are presented several studies, by category of the ICF, that evaluated ABI patients using wearable devices with the aim of estimating different clinical scores during the performance of activities of daily living.

2.3.1 Assessment of Activity

Research by Uswatte *et al.* involved one hundred sixty-nine stroke survivors using accelerometers in the upper-limb outside the laboratory for three days before and after therapy. The study also included completing the Actual Amount of Use Test (AAUT), an observational measure of the use of the more-impaired arm, and the Motor Activity Log (MAL), a patient-reported assessment of the use of the

more-impaired arm in daily life. Overall, it was shown that it was possible to extract clinically useful information about the upper-limb motor state using accelerometer data [52]. For the baseline, the correlation between the ratio of more-impaired to less-impaired arm regarding accelerometer recordings vs. AAUT score was 0.60, and vs. MAL scores were 0.52, respectively. Additionally, it was possible to point out that the duration of impaired-arm movement divided by the duration of unimpaired-arm movement showed to be a reliable and valid motor outcome.

Moreover, Barth *et al.* in 2020 analyzed the relationships between accelerometry and general compensatory movements of the upper-limb after stroke, and the translation of this measurement tool into clinical practice [36]. The data sources used included a compensatory movement score derived from video analysis of the Action Research Arm Test (ARAT) and accelerometer variables calculated from the same time point during study participation. The accelerometer provided measurement of activity counts that were used as an index of upper-limb use. The results indicated that the paretic upper-limb demonstrated compensatory movements. This was assessed with the ARAT scale and thus observing the compensatory movement score and comparing it with each accelerometer variable. This research presented that accelerometry is a powerful tool to measure movement quantity and identify general compensatory movement patterns of the upper-limb in ABI survivors. Essentially, higher movement while performing ADL, concerning time and variability, is indicative that the patient is expected to have fewer movement compensations and closer to healthy movement patterns. Thus, patients who don't usually incorporate their paretic limb in ADL tend to have the compensation of movement at the joints in the paretic limb. Therefore, the movement patterns are distant from the standard reference. This study reveals that there is also potential to derive new outcomes about the quality of movements from accelerometer data.

Patel *et al.*, focused on deriving FAS from tasks selected from the WMFT [53]. Twenty-four subjects participated in the study, clinicians reported outcome measures for fifteen tasks of WMFT, and six accelerometers data were collected for eight tasks of the WMFT. The data analysis included task segmentation, feature extraction, and feature selection (using WEKA implementation of the ReliefF algorithm), where features were ranked based on their importance. It was noted that features from distal segments were extracted more frequently than those deriving from proximal segments. Then, a Random Forest algorithm was used to estimate the total FAS score (i.e., the sum of individual scores for each of the fifteen motor tasks of the FAS). The results exhibited a relative RMS error for estimating the total FAS score of 5.76% when the eight tasks were considered. The best result was achieved when using only four tasks enabling a relative RMS error of 9.12%.

A recent study by Adans-Dester *et al.* assessed ABI patients while performing eight reach-manipulation functional tasks (Figure 2.6) from the WMFT, in a clinic setting [7]. The patients recruited included 16 stroke survivors and 21 TBI survivors, and all presented with residual upper-limb hemiparesis. The sensors used were seven accelerometers placed on the chest, arm and wrist, and on the index and thumb of the affected upper-limb. The approach consisted of generating clinical scores of: quality of movement, using the Functional Ability Scale (FAS); the severity of motor impairments, using the upper-limb Fugl-Meyer Assessment (FMA) (as will be seen in the next subsection).

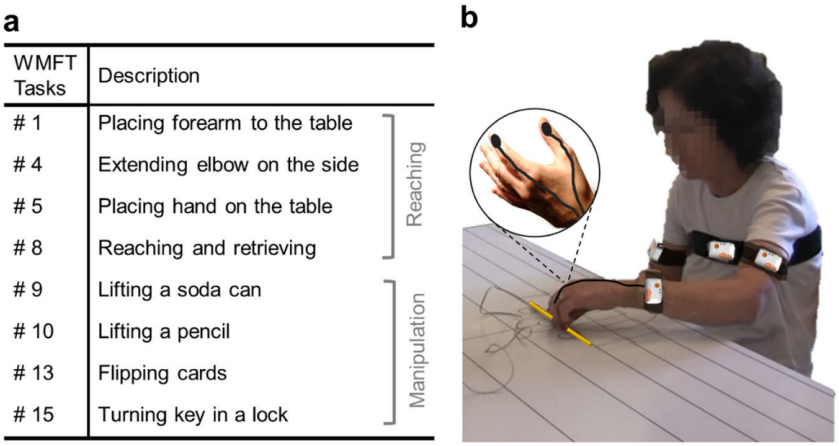


Figure 2.6: a) WMFT, functional tasks performed during the data collections b) position of the wearable sensors. Reproduced from [7].

The data analysis consisted of using the accelerometer data and applying an algorithm that existed to estimate the Functional Ability Score. The process of data analysis was similar to the previous study mentioned: first perform task segmentation, pre-processing of the sensor data (data was low-pass filtered to remove high-frequency noise and high-pass filtered to isolate the acceleration components due to postural adjustments), then feature extraction and selection (achieved using a correlation-based algorithm). Finally, it was possible to derive estimates of the quality of movement with FAS using Random Forest algorithms (Figure 2.7). The coefficient of determination for upper-limb score estimates was 0.79 for movement quality with the FAS.

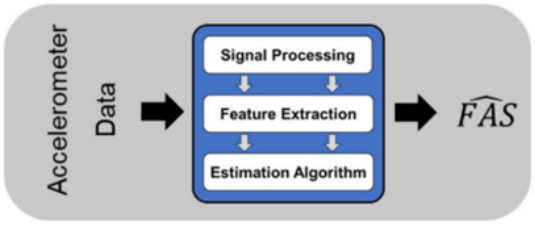


Figure 2.7: Proposed technique to estimate movement quality (FAS) clinical score. Reproduced from [7].

2.3.2 Assessment of Body Structure and Functioning

In order to do an automatic estimation of Fugl-Meyer scores, Del Din *et al.* assessed twenty-four stroke patients using six accelerometers connected to the hemiparetic side of the upper-limb: hand, arm, and trunk (Figure 2.8) [8]. The data used was collected in a previous-mentioned study by Patel *et al.*, which aimed to derive FAS from tasks selected from the WMFT [53]. The dataset also included clinician outcome measures for fifteen tasks of WMFT; however, sensor data was collected for only eight tasks of the WMFT. Thus, with the same dataset, the researchers now estimated FMA scores. To this end, it was performed feature extraction and selection (to optimize the feature set for each task) done in the WEKA environment using a ReliefF algorithm, and finally, it was proposed a Random Forest based algorithm [8]. The results were indicative that it was possible to estimate Fugl-Meyer Assessment scores by feeding a Random Forest with features derived from accelerometers during the performance of as few as a single item of the WMFT. Overall, the model proposed enabled the achievement of a root mean squared error as low as 4.7 points of the FMA. Nevertheless, a larger cohort of patients should be considered, and further

study of this technology is necessary—for example, estimation of sections of the FMA scores to better understand each section's movement biomarkers.

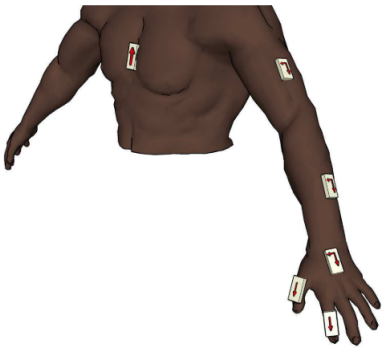


Figure 2.8: Scheme of sensors positioning and axes orientation to gather accelerometer data from stroke survivors during performance of tasks. Reproduced from [8].

Wang *et al.* developed research that proposed a Support Vector Regression (SVR) model to estimate FMA scores [54]. The study included twenty-four stroke patients performing four tasks while using two accelerometers (upper-arm and forearm). Clinicians assessed the patient, and data was recorded during the performance of the tasks. Once again, there was the need to pre-process data and extract features. Then, the ReliefF-SVR system was used to select the features. Ultimately, the ReliefF algorithm was implemented on WEKA for ranking features according to importance, and later SVR was used with the first N features. Finally, to determine the optimal number of features, which ended up being fourteen, numerous SVR models with a different number of features were built. The proposed model achieved a cross-validation prediction error value of 2.1273. Nonetheless, in future studies, it is desirable to also asses wrist-hand motor function and estimate the FMA score accordingly to have a complete assessment of the upper-limb motor function, which was not included in this paper.

A study by Yu *et al.*, aimed to remotely assess the motor function of twenty-four stroke patients in home settings (i.e., non-clinical) [9]. The study involved using two accelerometers and seven flex sensors to acquire data on the movements of the upper-limb, wrist, and fingers (Figure 2.9). Data were acquired from patients performing seven training exercises that represent the upper-limb. The study was designed in two phases. Firstly a clinical phase, where twenty-four patients were evaluated by two clinicians and also using the sensor setup. Then, a home setting phase, where solely five out of the twenty-four patients participated.

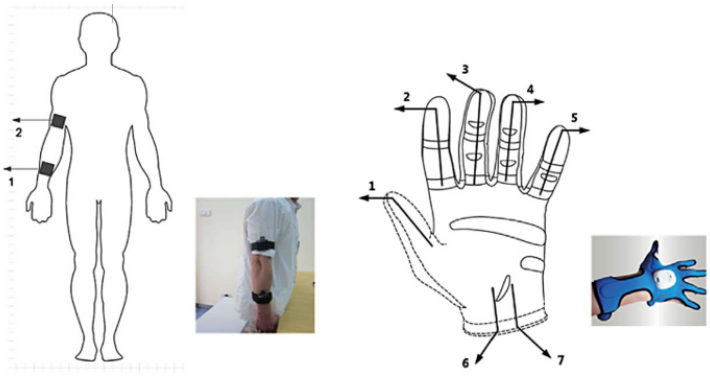


Figure 2.9: Sensor placements on the human body: (left) placement of two accelerometer sensors; (right) placement of seven flex sensors. Reproduced from [9].

Moreover, regarding the study by Yu *et al.*, the data analysis included the processing of sensor data, and they used an extreme learning machine (ELM) algorithm to derive the FMA estimates (Figure 2.10). As depicted in Figure 2.10, the algorithm had the goal of mapping the sensor data to clinical FMA scores, thus to this aim, first is necessary to pre-process the accelerometer and flex sensor data and then perform feature selection and extraction. The proposed FMA model is mentioned as being able to reach a coefficient of determination for upper-limb score estimates that are as high as 0.917.

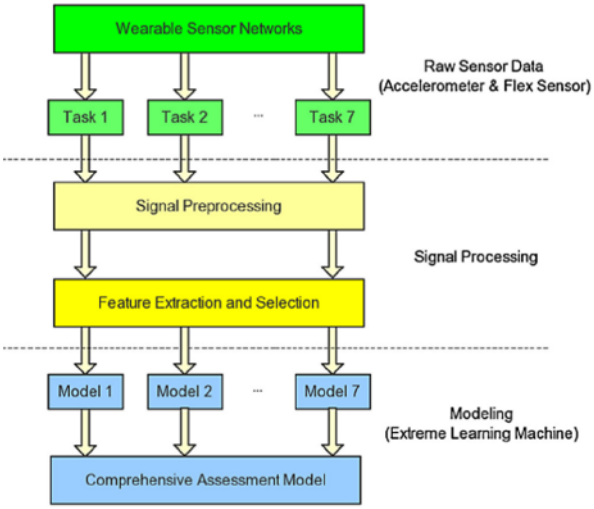


Figure 2.10: Flowchart of quantitative Fugl-Meyer assessment (FMA) model. Reproduced from [9].

However, this study only included impairment scores with FMA; it would be interesting to have other clinical scales involved that could provide information on the quality of movement. Another thing to consider is that flex sensors have shown to have major disadvantages such as the cost to build a control system for both hands, being fragile, and errors happening when the sensor is used for a long time due to changes in the sensor's flexibility [55]. Furthermore, the use of a glove may be a limitation for ABI patients while performing ADL. Therefore, the sole use of accelerometers to acquire data on the upper-limb's movements, wrist, and fingers would be advantageous.

A previously mentioned study by Adans-Dester *et al.* assessed ABI patients while performing eight reach-manipulation functional tasks (Figure 2.6) from the WMFT, in a clinic setting [7]. The approach consisted not only in deriving clinical scores of quality of movement with FAS but also the severity of motor impairments, using the upper-limb Fugl-Meyer Assessment (FMA).

The researchers tried different implementations of algorithms to estimate FMA scores. The best results were obtained when the estimated FAS was used in another module as input to derive the FMA score (Figure 2.11).

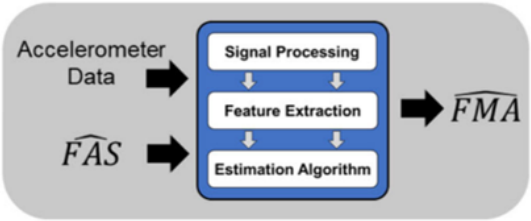


Figure 2.11: Proposed technique to estimate motor impairment (FMA) clinical score. Reproduced from [7].

The process of data analysis was the same as previously mentioned. Overall, it was possible to derive estimates of the quality of movement with FAS and impairment scores with FMA using Random Forest algorithms. Therefore, the proposed novel technique includes FAS score estimates (derived from the wearable sensor data) as supplementary input to the Random Forest method. The coefficient of determination for upper-limb score estimates was 0.86 for impairment severity.

2.4 Final Considerations

Overall, the development of ML algorithms depends on what is desirable to estimate. All studies mentioned used supervised ML techniques because most of these studies have clinical evaluations that are done while collecting data. Therefore, it is relevant to use supervised algorithms because labels are useful from the clinician's point of view. Unsupervised ML techniques are more used when it is desirable to find structures in unlabeled data.

Usually, the standard kinematic measures that researchers use to analyze participants' bimanual object manipulation tasks and reaching performance are the movement time (i.e., time period from movement onset to movement offset), spectral arc length (i.e., dimensionless measure of the arc length of the Fourier magnitude spectrum of the velocity signal), peak velocity (i.e., highest point on the resultant velocity curve), peak acceleration (i.e., highest point on the resulting acceleration curve), mean velocity (i.e., average of the resultant velocity signal), and mean acceleration (i.e., average of the resulting acceleration signal) [37]. Generally, there is a noted higher correlation with movement time, spectral arc length, and peak velocity related to the Fugl-Meyer assessment for upper-extremity score [37].

By analyzing different articles, it was possible to enumerate some advantages of the use of accelerometers over classical assessment techniques for ABI rehabilitation management, namely: the non-invasiveness, flexible application promoting an easy assessment of the paretic upper-limb in the clinic, and comfort. However, sensor costs and further miniaturization would be necessary to enable a more unconstrained environment when wearing multiple sensors over hours. Additionally, these instruments are tools to assess motor function more objectively, having the potential to predict clinical assessments scores estimation.

To sum up, these active monitoring systems have user requirements such as accuracy, comfort, setting (home and clinic), scalability, low-cost [5], ease-of-use, and automation improvements. These systems also encompass wearability factors, such as weight, placement, attachment, material, and sizing (e.g., MEMS have facilitated the development of miniaturized accelerometers that are important for finger movement monitoring.).

Currently, accelerometers have the potential to provide high kinematic accuracy measurements of ABI patients performing ADL. The ML algorithms can be developed based on the different scales to be assessed, sensor placements, and other clinical parameters that would be of interest. Thus, these devices enable a transition from clinician-reported outcomes to sensor-based outcome estimates. Hence, by tracking motor recovery, it is possible to evaluate the patient's response to an intervention. Consequently, allowing clinicians to change the treatment plan that will hopefully lead to maximizing acquired brain injury patients' motor gains.

Chapter 3

Materials and Methods

This chapter addresses a detailed description of the materials and methodology followed in this dissertation. Two main subsections are presented:

a) Laboratory Data Collection: Clinical information on the participant's demographics and inclusion criteria as well as the study design are described. Following, experimental procedures and technical specifications regarding the IMU operation and the motor task performance are explained. Finally, the clinical scales used to assess the patient's motor capabilities are enumerated.

b) Machine Learning Algorithm: The specific techniques and models used to analyze and derive the clinical scores are expressed, ranging from the signal processing techniques to the feature extraction and selection and, ultimately, the ensembled ML algorithm.

3.1 Laboratory Data Collection

3.1.1 Preliminary Procedures

The experiment was carried out in the Motion Analysis Laboratory (MAL), Department of Physical Medicine and Rehabilitation of Harvard Medical School at Spaulding Rehabilitation Hospital in Boston. The study was funded by the National Institutes of Health (NIH) Award #R24HD065688: "Improving Outcome Measurement for Medical Rehabilitation Clinical Trials" and had a data collection duration of 25 months. This Dissertation is part of this study, and all the analyses have been accomplished on previously collected data. The Institutional Review Board (IRB) at Spaulding Rehabilitation Hospital approved this study. All study participants or their legally authorized delegates were required to sign an informed consent form, which described the study's objectives and the associated health risks and benefits.

3.1.2 Participants

The study was conducted on a heterogeneous sample of thirty-seven ABI individuals with upper-limb hemiparesis. Of these patients, sixteen were stroke survivors, while twenty-one were traumatic brain injury survivors. Participants' characteristics are summarized in Table 3.1.

Subjects were assigned a study number, which was used for documentation and analysis throughout this dissertation. Due to this study's nature, tracking changes in the level of motor impairment is key. Thus, to be eligible for the study, the subjects had to meet the following criteria:

Table 3.1: Demographic information related to the participants' (N=37) clinical characteristics.

Clinical Demographics	
Number of participants (n)	37
Stroke	16
TBI	21
Age (years)	42.6 ± 19.0
Gender, n (%)	
Male	26 (70.3)
Female	11 (29.7)
Handedness, n (%)	
Left	3 (8.1)
Right	34 (91.9)
Affected Side, n (%)	
Left	17 (45.9)
Right	20 (54.1)
Chronicity (months)	8.3 ± 14.9
Inpatient/Outpatient, n (%)	
Inpatient	30 (81.1)
Outpatient	7 (18.9)
Days Between Assessments (days)	36.2 ± 23.6
Total Hours of Therapy (h)	16.0 ± 9.8
Mini-Mental State Examination (score)	25.5 ± 5.0

Inclusion criteria

- Sustained a unilateral stroke, both ischaemic and hemorrhagic or a focal traumatic brain injury (including scalp injury, skull fracture, and surface contusions, generally caused by contact)
- Male and female, age range 18-70 years old
- Must have chosen to participate in a rehabilitation program (inpatient or outpatient) or research study involving upper extremity therapy lasting at least 6 weeks
- Reached a score of 15-55 (moderate-severe) out of 66 points, evaluated through the Fugl-Meyer Assessment for Upper Extremity (FMA-UE)

Exclusion criteria

- Participants were assessed on their ability to follow instructions using the Mini-Mental State Examination (MMSE). Participants were excluded if there were severe cognitive impairment (MMSE < 10) that could affect the ability to understand instructions for motor tasks and assessment tools.

3.1.3 Study Design

The experimental data collection consisted of two separate laboratory visits, namely, the first visit being a Baseline assessment and the second, a Post-Treatment assessment. At Baseline, data was acquired before the patient initiated a rehabilitation program; at Post-Treatment, data was recorded at the end of the rehabilitation program (Figure 3.1). The rehabilitation program consisted of an inpatient or outpatient upper-extremity motor rehabilitation intervention at Spaulding Rehabilitation Hospital, lasting at least six weeks.

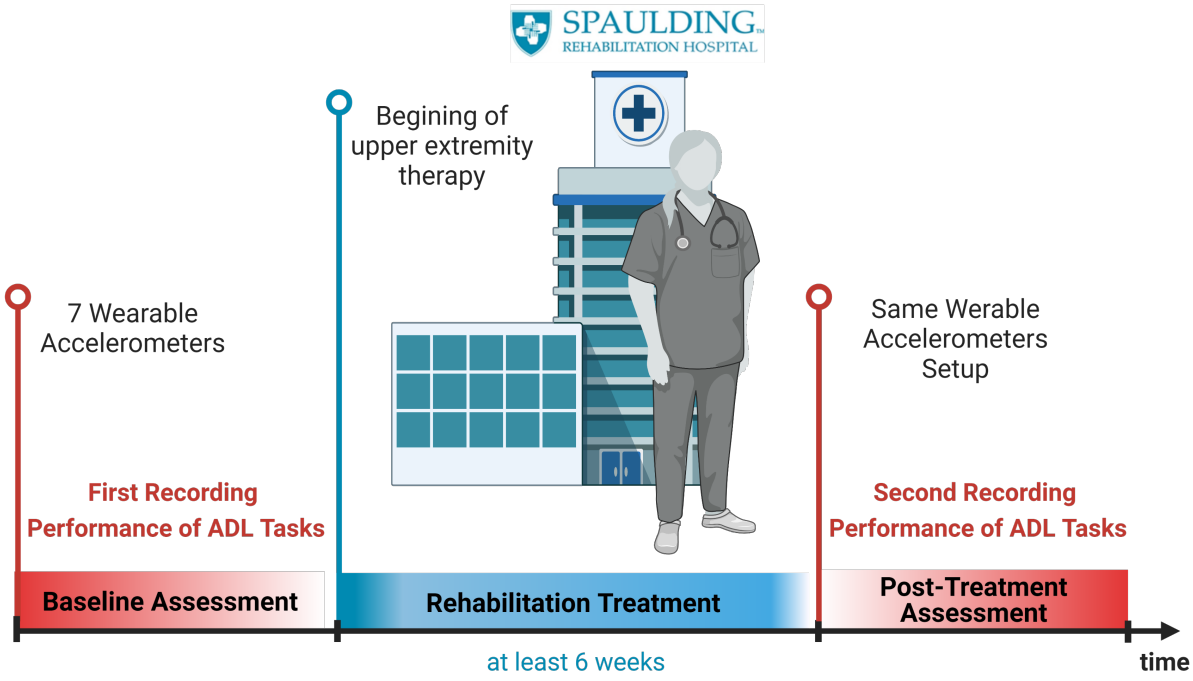


Figure 3.1: Study Design Timeline. The laboratory assessment visits included two data recording sessions during the performance of ADL tasks: Session 1, Baseline; Session 2, Post-Treatment follow-up. The assessments are highlighted in red in the timeline. The rehabilitation treatment, either inpatient or outpatient, is stressed in blue in the timeline.

3.1.4 Experimental Procedures

Moving on now to consider experimental procedures, which were repeated during Baseline and Post-Treatment assessment visits.

Sensor Setup

To enable the assessment of the effects of the upper-limb rehabilitation interventions, subjects were instrumented with a total of seven wireless IMUs (Shimmer2r by Shimmer Sensing, Dublin), respectively, five 3-axis main unit accelerometers (Figure 3.2) and two 2-axis accelerometers.

Figure 3.2: Shimmer2r IMU module.

The specifications of the Shimmer2r sensors are shown in Table 3.2. As was pointed out, the data acquisition was performed using the IMU option of integrated accelerometers, capturing data at a sampling frequency of 51.2 Hz. These sensors are known to be small, compact, ultra-low power, and lightweight wearable devices that enable the patient's full mobility during a motor performance, having the advantage of providing us researchers with highly accurate and reliable raw data with low noise.

Table 3.2: Specifications employed with the Shimmer2r accelerometers.

Shimmer2r Unit Specifications	
Weight	27 g
Dimensions	53 mm x 32 mm x 19 mm
Processing	MSP 430 microcontroller (8 mHz,16 bit)
Storage	Integrated 2GB microSD card slot
Battery	450 rechargeable Li-ion
Integrated Axis Acc	Freescale MMA7361
Acc Range	$\pm 1.5g$
Sampling frequency	51.2 Hz

The accelerometers were positioned in the upper-limbs as depicted in Figure 3.3, respectively the 3-axis accelerometers were placed on the chest (sternum height), upper-arms (mid-biceps, frontal), wrists (above radius and cubitus styloid, dorsal), and the 2-axis accelerometers were set on the index and thumb fingers (dorsal part of the distal phalange) of the affected, hemiparetic, side. The two additional sensors on the fingers of the hemiparetic side of the patient aim to track patient motor-specific patterns, primarily in fine specialized motor control performance of tasks.

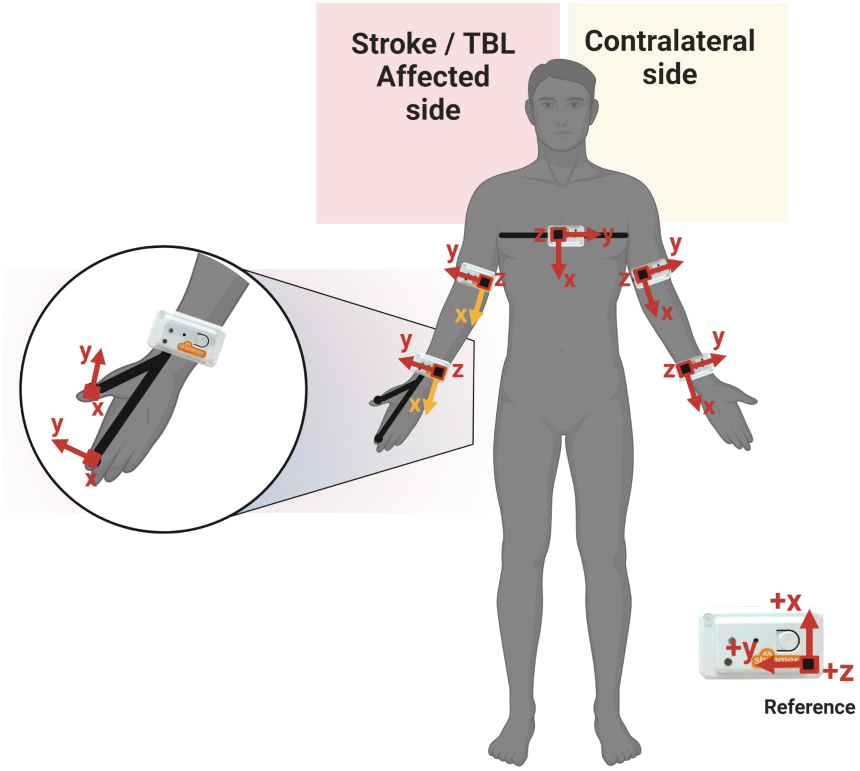


Figure 3.3: Shimmer2r IMU modules connected to the Shimmer baseboard and placed on the affected and contralateral side of the patient. The reference Shimmer2r coordinate system is as follows: X-axis, vertically up in the sagittal plane; Y-axis, leftward in the coronal plane; Z-axis, pointing outward on the horizontal plane. The yellow-colored arrows on the patient's affected side mean that an axis inversion algorithm was developed in order to match the axis orientation of the contralateral side. Zoomed are the 2-axis accelerometers connected to the main wrist unit of the patient's affected hemiparetic side.

Motor Tasks Performance (ADL)

After the sensor setup was complete, participants were guided to perform a series of eight standardized tasks specifically chosen to reproduce motor patterns of activities of daily living. Eight of the 15 tasks of the WMFT were selected and adapted, as previous studies indicated that similar subsets could enable researchers to accurately estimate the total FAS score [53].

The tasks include bimanual or unimanual manipulation of objects or both and fine or gross motor control or both. Figure 3.4 depicts the eight tasks; participants were video recorded while performing different repetitions of each task, and a capacitive sensor was used to mark the beginning and end of each task repetition.

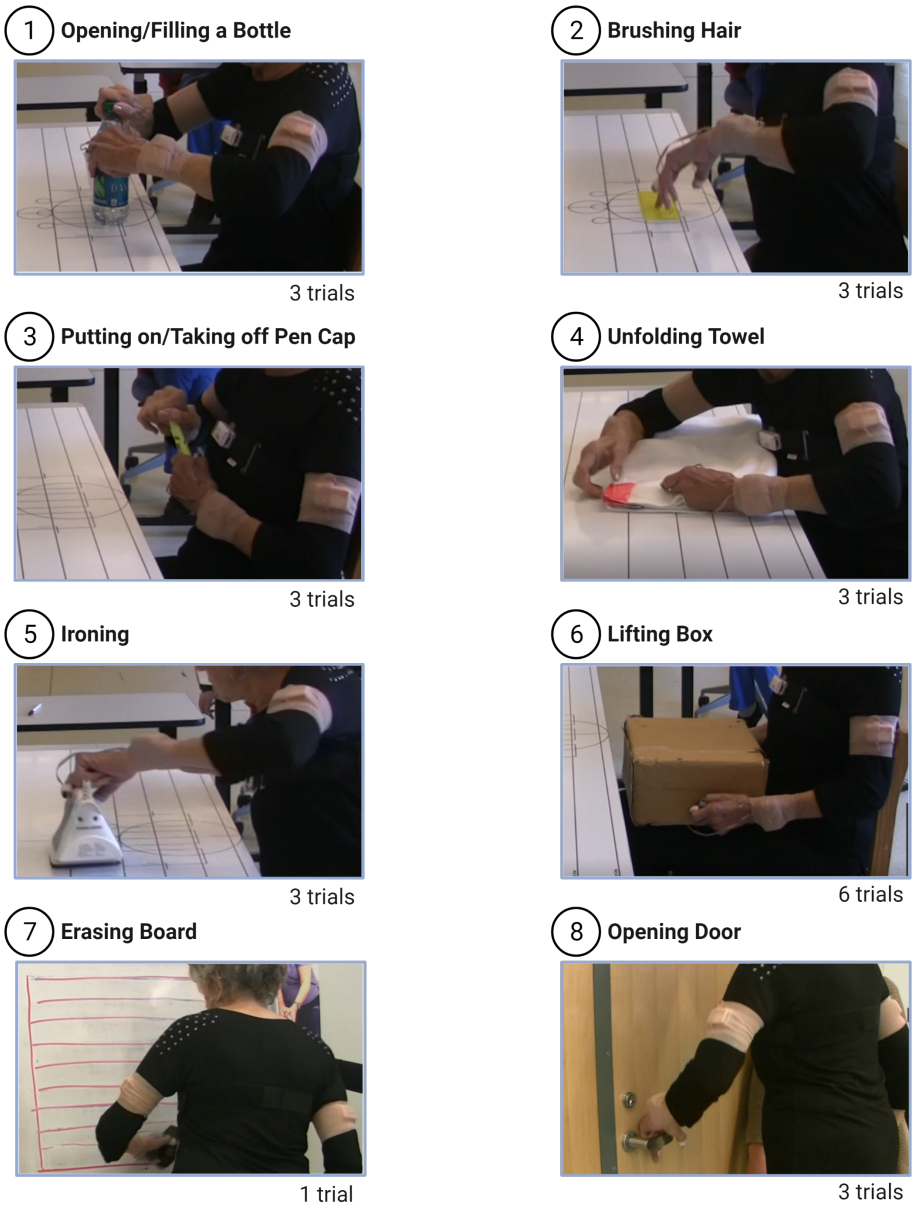


Figure 3.4: Activities of Daily Living tasks performance in the MAL laboratory setting. The eight tasks are enumerated and exemplified by a deidentified subject. The number of trials repeated per task is described. The IMU sensors were placed on the upper-limb using a self-adherent wrap.

Clinical Assessments

A study therapist or a clinician conducted clinical tests to assess the functionality of the upper extremity before and after the rehabilitation intervention program. The clinical evaluations included:

- Fugl-Meyer Assessment for Upper Extremity: providing a measure (a score from 0 to 66) of upper extremity motor impairment, including coordination and speed, assessed per subject at Baseline and Post-Treatment.
- Functional Ability Scale: measuring the quality of the movement (a score from 0 to 5), assessed per trial of the task being performed at Baseline and Post-Treatment.

3.2 Machine Learning Algorithm

The ML algorithm's development had the ultimate goal of deriving FAS (see Appendix B) and FMA (see Appendix C) clinical scores from data obtained in the clinical assessments. The workflow of the summarized steps followed in the ML algorithm to estimate the FAS and FMA are depicted in Figure 3.5. The routines to develop the algorithm were implemented using MATLAB R2021a software (The Mathworks Inc., Natick, MA, USA) [56].

A detailed explanation concerning the steps involved in the algorithm is described in the following subsections. I will start by explaining the pre-processing methods, followed by the feature extraction, feature selection, and the model used to train the features, validate the model and estimate the FAS and FMA. Note that it was taken into account the patient's motor ability knowledge learned from the FAS estimation model and used the FAS model as further input to derive the severity of motor impairments with FMA.

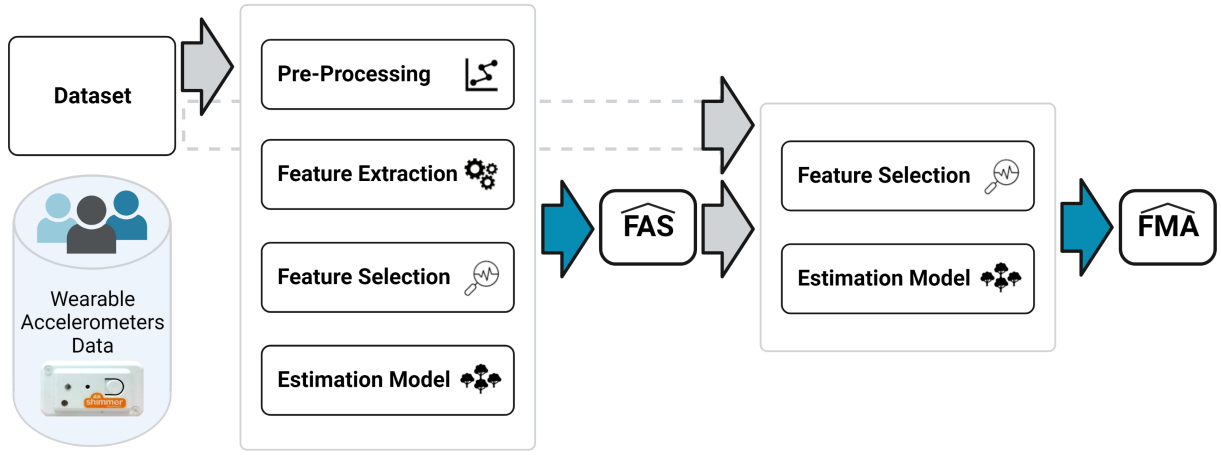


Figure 3.5: Workflow of the ML pipeline followed to estimate the clinical scores, FAS and FMA. Gray arrows indicate the model inputs, and blue arrows the outputs. Single task FAS predictions provided by the first model were used as additional features input of the second model in order to estimate the FMA.

3.2.1 Signal Pre-Processing

The available data was the raw output of the Shimmer2r. In this subsection, each axis (i.e., unit channel) signal was individually pre-processed with different methodologies to improve the signal quality and model performance of further stages (Figure 3.6).

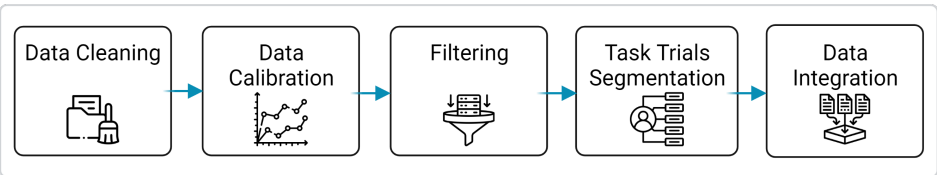


Figure 3.6: Pre-Processing Pipeline.

Data Cleaning

The first step of the data pre-processing was to identify the relevant Shimmer2r axis. The data for the hemiparetic wrist unit index and thumb sensors were acquired for 5 extra channels in total. Therefore, a method was employed to identify the 4 channels of the index and thumb accelerometers. To this effect, the correlations between the axis of the affected-side main wrist unit, consisting of a 3-axis from the wrist and a 5-axis (i.e., five extra channels) from the fingers sensors, were computed.

The final step of the data cleaning was to fill in missing values and correct inconsistencies in the data. Such as invalid scores and missing data, including: missing raw data of sensor's .csv files; trials that had the wrong structures due to invalid markers (start/stop time to trim the trials); trials with an invalid FAS score; trials with invalid channels data.

Data Calibration

As mentioned, the Shimmer2r units provide raw data as output, in this case uncalibrated. Hence, as the next step, an algorithm was developed to calibrate the raw data to the SI units of acceleration. Since there was missing information regarding the IDs and data acquisition calibration matrices of the sensors, a specific method to calibrate the data had to be developed. Calibration is relevant since, during the data collection, the sensor's position changed across subject assessments. Hence, the calibration removes a confounding factor, being desirable to compensate for the offset.

The algorithm was developed to primarily detect a resting period in the full task time series; our hypothesis is that by identifying this window of interest, will be possible to derive the missing offset vector for a specific task needed to calibrate the data.

The calibration of the tri-axial inertial sensor signal to SI gravity units is achieved through the implementation of the formula:

$$c = R^{-1}K^{-1}(u - b) \tag{3.1}$$

where c = 3x1 calibrated signal vector; R = 3x3 alignment matrix; K = 3x3 sensitivity matrix; u = 3x1 uncalibrated signal vector; b = 3x1 offset vector.

The methodology to find b, the offset vector, is described next.

The sensor's signal for a given task was segmented into small windows of 1 second. Then, the individual windows are compared, where specific criteria were defined to select the more stable (i.e., resting) window. In this case, the criteria that met our requirements was to find the window with the

lowest root-mean-square (RMS) value of the magnitude of the signal (magnitude time-series defined as the square root of the components' squares for each data sample).

Once the window of interest (flattest magnitude signal) is found, the algorithm to search for the best-offset vector in the window of interest is applied. A search method was used for this purpose.

The default offset vector, $b\_original$, for the sensor is known and provided by the manufacturer specifications; regardless, this vector doesn't correspond to the actual offset of the sensor at data acquisition. Still, this vector is used as a framework to search for the optimal solution. A temporary offset vector, b_temp, is iteratively computed by the sum of the $b\_original$ to the search range. Thus, within the window of interest of the uncalibrated acceleration, it was computed the new temporary calibrated vectors, c_temp, given by (Equation 3.1), then with c_temp the temporary magnitude, mag_temp, was calculated.

The expected magnitude acceleration, mag_rest at rest is 9.81 m/s^2. Accordingly, a cost-driven loop iterates the search range until the offset matrix, b_temp, that minimizes the cost is found. The cost was defined as the absolute difference between mag_rest and mag_temp.

Finally, after discovering five optimal offset vectors for each of the main unit sensors, Equation 3.1 is applied to all data across tasks in those sensor positions. An assumption is made that the sensor positions were unchanged during the performance of the different tasks.

Filtering

The calibrated time series for each axis of the accelerometer units, sampled at 51.2 Hz, have been filtered with a Butterworth passband filter with the following specifics: a filter order set to 6 for computation cost reduction purposes and a frequency passband between 0.5 and 10 Hz.

Figure 3.7 represents the Bode diagram which describes the frequency response of the linear time-invariant system. The filter has the advantages of presenting a linear phase response in the passband and a flat band response (i.e., less ripple); nevertheless, it encompasses a slow cutoff.

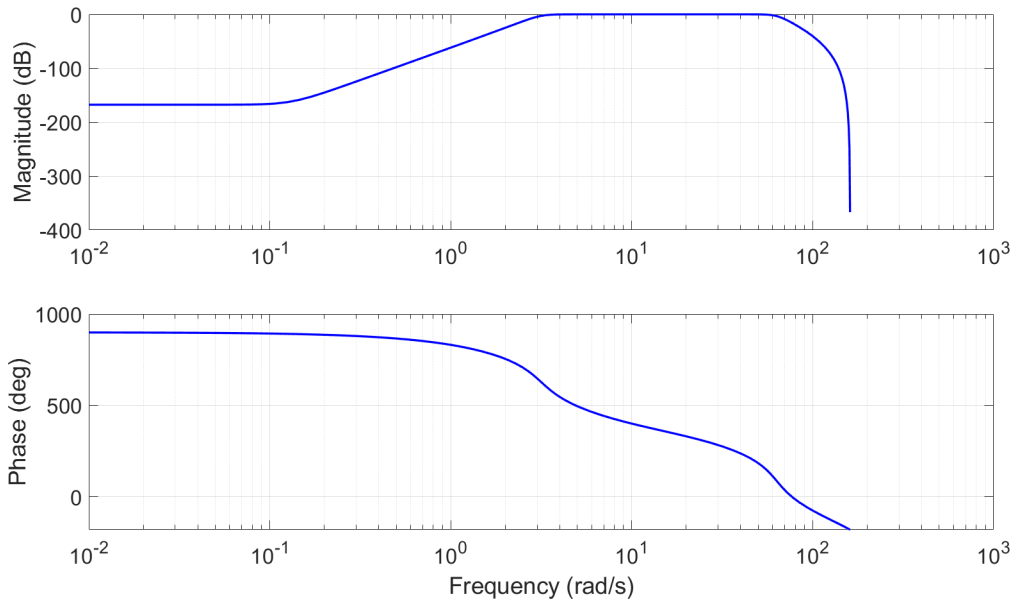


Figure 3.7: Bode diagram of the 6th order Butterworth filter with a passband of [0.5-10] Hz. Semi-log superior plot represents the gain (magnitude) in dB and the inferior plot shows the phase in degrees.

These criteria were specified since upper-limb movements are highly variable and specific to each patient. Notably, the literature indicates that the motion frequency of the upper-limbs of humans is usually lower than 10 Hz [57], [58], [59]. Likewise, a 0.5 Hz cutoff is usually relevant to minimize the impact of the orientation of the sensors that may differ among the subjects, limit the impact of postural adjustments, and attenuate the low-frequency integration drift characteristic of inertial systems that can encompass acceleration measurement errors which when integrated will conduct to progressively more significant errors [59], [60].

Segmentation and Integration

The last step in the pre-processing was to segment the task data into trials and integrate the data where each task trial is stored individually, considering the affected and contralateral sides sensors.

Using the time-stamp information provided by a long press of the capacitive sensor during the recordings, the beginning and end markers of the trials have been tracked. Thus, it was possible to automatically segment the accelerometer signals into the desirable trials.

Additionally, as depicted previously, in Figure 3.3, at this stage of the pre-processing, it was necessary to invert the X-axis signal of the affected side upper-arm and wrist sensor units in order to match the axis orientation of the contralateral side. Furthermore, in unimanual tasks 2, 5, 7, and 8, it was only considered the data from the trials performed with the affected side, which provide the motor patterns that are of interest for capturing in the clinical scores.

The integration of the data is essential to further analysis. Therefore, the organization was conducted in terms of task trials. Each trial data is stored separately from the others with the respective FAS score, grouping all the patients but maintaining the stroke or TBI difference. In the following subsection, each trial was analyzed to extract features.

3.2.2 Feature Extraction

The primary goal was to compute features that could recognize the leading characteristics of patient-specific upper-limb gestures, which are challenging to identify in general, particularly in atypical populations such as stroke and TBI.

As it is known, many of these tasks don't have a unique standard way to be performed and accomplished, and this should be something that is possible to capture with standard features of acceleration like speed, smoothness, and coordination.

There have been several studies that showed the relevance of certain features to identifying functional capability of patients [53], [7], [61], [62]. As a result, and after analyzing our patients' video recordings, it was possible to further engineer new feature categories directly associated with the movement of the patients, as presented in Table 3.3. For instance, it was noticed that patients with lower scores who tend to have lower performance had a pattern of a jerky movement, and so this is just something that was checked and considered. For patients with higher scores, it was remarked that, for example, the movement was smoother, and so there were fewer peaks in the signal. On the other side, features concerning sensor correlation are important, especially in tasks involving bilateral contributions for both arms, where it is desirable to check if the time series readings exhibit similarity [63].

Table 3.3: Engineered data feature macrogroup and descriptions. The features were extracted from the wearable sensors units calibrated and filtered time series data.

Data Feature Macrogroup	Description
Mean	Arithmetic; Harmonic; Interquartile
Spread	RMS; Interquartile range; Absolute deviation
Smoothness	Distance to filtered signal
Jerk	Normalized by velocity; RMS; Maximum frequency magnitude
Power Spectrum	Dominant frequency; Ratio between total energy and energy in secondary peaks; Energy of secondary peak; Energy in dominant frequency; Energy around dominant frequency
Entropy	Signal entropy
Speed	Maximum; RMS; Mean
Autocovariance	Range of autocovariance
Skewness	Signal skewness
Kurtosis	Signal kurtosis
Magnitude	Magnitude per sensor; Range magnitude acceleration; Range magnitude speed; Range magnitude displacement; Maximum; RMS mean and standard deviation; Entropy of acceleration; Maximum frequency of magnitude
Correlations	Between pairs of sensors time series; Between pairs of sensors magnitude

3.2.3 Feature Selection

After extracting the features, the next step was to perform feature selection, which is crucial to identify the optimal subset of engineered features that better capture the upper-limb movement characteristics observed. However, the dataset is imbalanced, meaning that the clinical scores are unevenly distributed and highly dependable on the task being performed; thus, for some clinical scores ranges, there are few data points.

Dataset Rebalancing

Accordingly, after analyzing the robust methods available to rebalance the dataset, and due to the success of recent synthetic methods, such as SMOTE, it was opted to use a synthetic point generation method that resampled the data to balance it and is known to reduce the bias and use adaptively learning. The technique is ADASYN, an Adaptive Synthetic sampling approach [64] and can be briefly described as follows:

The input is the training dataset D_{tr} with m samples $x_i, y_i, i = 1, ..., m$ where x_i is an instance in n dimensional feature space $\mathbf{X}$ and $y_s \in Y = \{1, 2, 3, 4, 5\}$ is the class label with the largest number of points. Define m_s and m_l as respectively, the number of minority and majority class examples. Therefore, $m_s \leq m_l$ and $m_s + m_l = m$. Then, the procedure is:

(1) Compute the degree of class imbalance with:

$$d_c = m_s/m_l \text{ where } d \in [0, 1] \tag{3.2}$$

(2) If $d_c < d_{th}$ then (d_{th} is a preset threshold for the maximum tolerated degree of class imbalance

ratio):

(2.1) Then, the next step would be to calculate the number of synthetic data points that need to be generated for the minority class c:

$$G_c = (m_l - m_s) \times \beta \tag{3.3}$$

Where $\beta \in [0, 1]$ is a parameter used to specify the desired balance level after the generation of the synthetic data. $\beta = 1$ means a fully balanced dataset is created after the generalization process.

(2.2) Following, per each example $x_i \in c$, find K nearest neighbors based on the Euclidean distance in *n* dimensional space, and calculate the ratio r_i defined as:

$$r_i = \Delta i/K, \; i = 1, ..., m_s \tag{3.4}$$

where Δi is the number of examples in the K nearest neighbors of x_i that belong to the majority class, therefore $r_i \in [0, 1]$.

(2.3) The next step would be to normalize r_i according to $\widehat{r_i} = r_i / \sum_{i=1}^{m_s} r_i$, so that $\widehat{r_i}$ is a density distribution ($\sum_i \widehat{r_i} = 1$)

(2.4) Calculate the number of synthetic data points that need to be generated for each minority example x_i:

$$g_i = \widehat{r_i} \times G_c \tag{3.5}$$

where G_c is the total number of synthetic data examples that need to be generated for the minority class c as defined in Equation (3.2).

(2.5) For each minority class data point x_i, generate g_i synthetic data examples.

This method was applied at two points: first, before carrying out the feature selection procedure, and then, for training, where a Leave-One-Subject-Out cross-validation (LOSOCV) approach was used to validate each model's performance. The LOSOCV works by iteratively leaving out one subject's data as the validation dataset and using the remaining subjects as training data for our model.

The characteristics of the ADASYN model that were found advantageous are the fact that it provides a good generalization, and at each LOSOCV iteration, data is synthesized without information about the patient the prediction is completed on, thus avoiding overfitting.

After rebalancing the dataset, all features have been normalized to have zero mean and unity variance. For this purpose, the function $zscore$ [65] was employed to normalize the dataset. Afterward, the data cloud point distribution of synthetic and real points was observed; here, the goal was to check if the synthesized points represented a good generalization of the real data points.

Feature Selection Methods

Feature selection is essential to improve the training performance by excluding those features that are not suitable for the prediction. This is especially relevant for relatively small datasets where the algorithm does not have a lot of observation points to detect exactly the boundaries of the data clusters. Thus, this technique dimensionality reduces the number of features to the ones which provide the best predictive power to model a set of data.

The basic framework for implementing feature selection is composed of two families: the wrappers and the filters. Usually, filters have the advantage of their simplicity and their ability to be implemented without requiring a lot of computational resources. On the other hand, a wrapper method uses a learning algorithm to improve its performance and takes into account the accuracy of a candidate feature subset.

For the scope of our analysis, four different supervised methods were selected, which use different metrics to compute the relevance of the features.

Filter Type:

- Minimum Redundancy Maximum Relevance (MRMR)
- ReliefF
- Neighborhood Component Analysis (NCA)

Wrapper Type:

- Random Forest built-in feature importance

This choice was based on the need for feature selection methods that were computationally efficient yet sensitive to complex association patterns, e.g., feature interactions.

The entire balanced, normalized dataset was used as input for feature selection. Following, the performance of the different methods was compared. The Random Forest [66] built-in feature importance was preferred in this dissertation, as will be explained further in the results section. Specific details on how Random Forest built-in feature importance operates will be completed in the subsequent methods subsection due to first being crucial to comprehending how the RF ensemble algorithm works.

3.2.4 Random Forest Ensemble Algorithm

To estimate the FAS and FMA clinical scores were utilized a Classification and Regression Treebagger, a flexible tool based on the Random Forest (RF) technique [66]. Our initial choice was based on two primary reasons. First, bagging has been shown to enhance the accuracy of random features by combining decisions from multiple decision trees. Secondly, it can provide ongoing estimates of the generalization error of the combined ensemble of trees. Furthermore, presented below are the main RF techniques advantages that meet our cohort and analysis requirements [67],[66],[68]:

1. Handles small datasets
2. Suitable for both Regression and Classification problems
3. Doesn't require a statistical distribution of the data
4. Accounts for interactions and nonlinearities between variables
5. Prediction with OOB data avoids overfitting
6. Allows for exploring a considerable number of distinct explanatory variables and only emphasizes those of high explanatory power at each node split

In MATLAB [56], the mentioned algorithm is executed by command $TreeBagger$ [69]. For this implementation, was analyzed the model performance with different hyperparameters. The final setup consisted of: a number of 100 trees, m_{trees}, grown on the most important features, n_{feats}, given by the feature selection; the minimum number of observations per tree leaf used was the default (1 for classification and 5 for regression). The inputs are the predictor variables (n_{feats}) and the class labels (FAS or FMA scores).

The basics of the process performed in the Bagging training RF algorithm are depicted in Figure 3.8 and can be described as follows [70], [71]:

(1) From the original training set, m_{trees} bootstrap (i.e, dataset sub-sampling) samples were taken.

(2) Each bootstrap sample will then grow an unpruned tree for regression or classification, using a random sample of the feature predictors n_{feats}.

(3) Then, a greedy algorithm is used to find the best-split points among the sampled n_{feats}, which can represent similarity in trees structure and high correlation in their results.

(4) Based on the ensemble of bagged decision trees, the m_{trees} predictions were aggregated (by majority votes for classification, and, average for regression problems) by using the function $predict$ [72]. The predictions are done on the validation data using the above-mentioned LOSOCV approach.

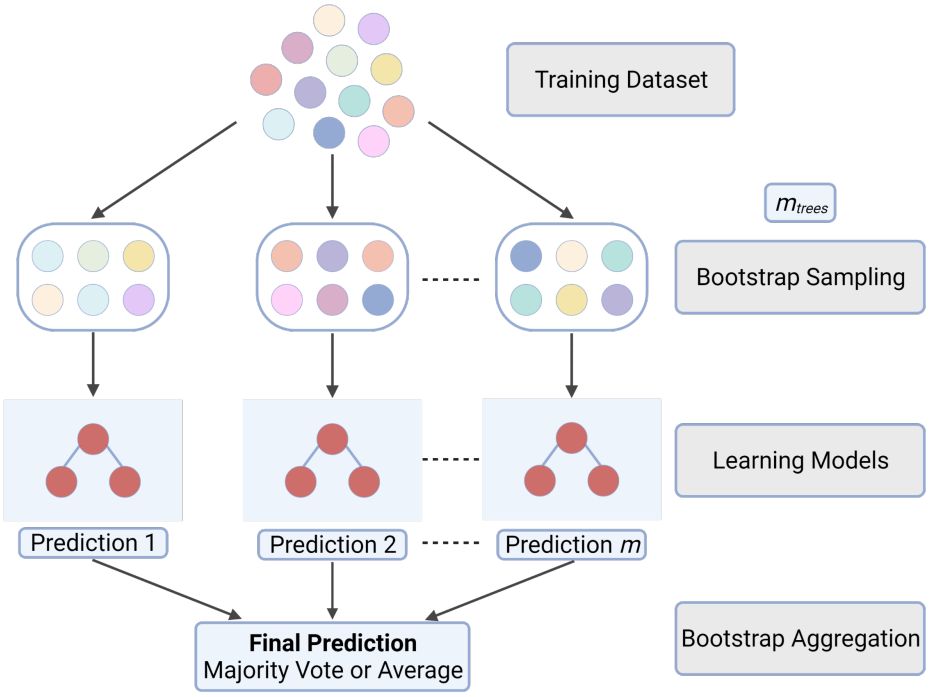


Figure 3.8: Schema of the Random Forest Emsemble algorithm.

Details on the Random Forest built-in feature importance

Understanding the methodology of a Random Forest Ensemble Algorithm was crucial to now describing and comprehending the details concerning the RF built-in feature importance method.

As the name indicates, this is a built-in method of feature importance selection of the RF. To check the optimal feature number, a RF algorithm is pre-trained on all features with $Treebagger$ [69].

The final optimal feature importance number is selected with the Out-of-bag (OOB) error. For each bagged tree, some instances may be sampled multiple times for any given predictor, while others may not be sampled at all (i.e., OOB instances, validation points) and are not seen by the predictor over training. Consequently, the evaluation of the ensemble is made on those instances by averaging (for regression) or considering the majority vote (for classification) of each predictor.

In the $Treebagger$ this error is given by the $OOBPermutedPredictorDeltaError$ parameter. Thus, the total errors of all the configurations are compared, and the configuration with the lowest $OOBPermutedPredictorDeltaError$ is selected. Accordingly, solely the most predictive features were used in our models as further inputs.

3.2.5 FAS and FMA Estimation

The algorithms developed to estimate the FAS and FMA can be broken down into two main points.

a) Model to Predict Single Tasks Scores: firstly, the goal was to estimate FAS and FMA clinical scores for the individual ADL tasks.

b) Model to Predict the Final Estimate of the Total Clinical scores: secondly, an aggregation of the FAS and FMA predictions was performed to derive the final estimate of the total clinical scores.

An overview of the algorithm developed for the FAS and FMA clinical scores is shown in Figure 3.9. For the task-specific RF models (100 trees), the inputs constitute the labels of the assessed clinical scores and the respective features concerning the wearable sensors data. The model outputs several single clinical scores predictions provided by tasks.

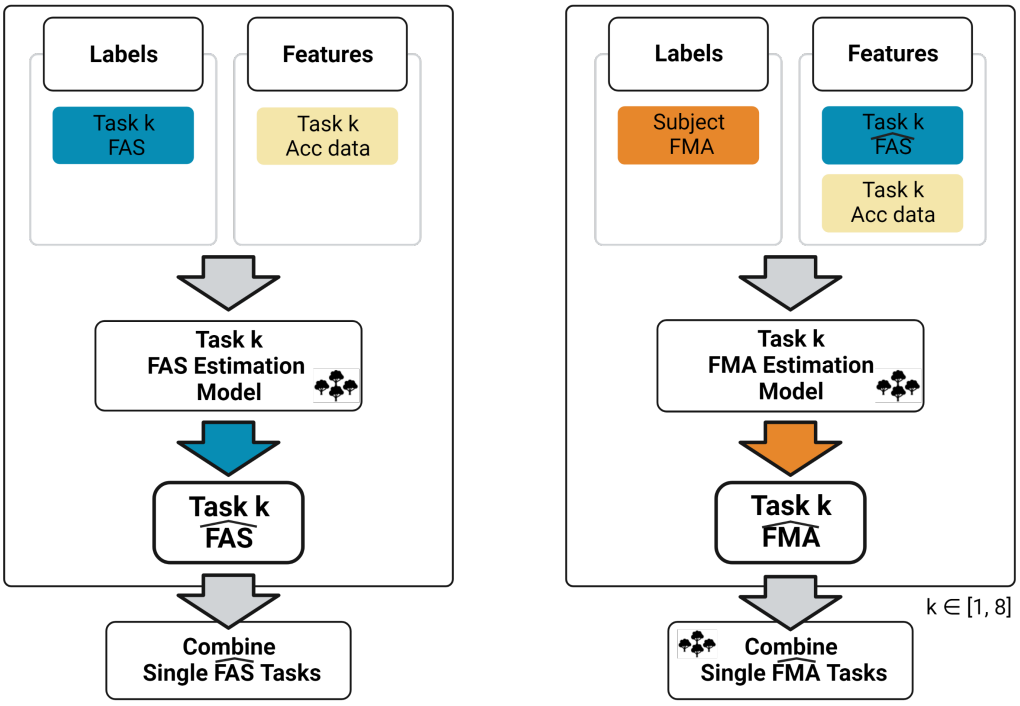


Figure 3.9: Algorithm overview for deriving the FAS and FMA. Gray arrows indicate the model inputs, and blue arrows the outputs. On the left, the algorithm to derive the FAS is shown. For this panel, the inputs are the FAS scores assessed per subject at the Baseline and Post-Treatment time points for each task trial with $k \in [1, 8]$. The features include wearable accelerometer data and are also given as input for the task-specific RF model. Ultimately, per subject and time point, a single FAS score prediction is the model's output. A model was used to combine the FAS predictions. On the right, is presented the algorithm to estimate the FMA with inputs the labels are the FMA scores assessed per subject at the Baseline and Post-Treatment time points. The features include the single tasks FAS predictions with $k \in [1, 8]$, and the wearable accelerometer data and are also given as input for the task-specific RF model. Ultimately, per subject and time point, a single FMA score prediction is the model's output. Another RF model was used to combine the single FMA predictions.

A total of four models were developed to predict the single FAS tasks trials scores:

a.1) FAS Model 1: The FAS scores from 1 to 5 were used as labels in a RF algorithm.

a.2) FAS Model 2: The FAS scores 1 and 2 were merged into one score, leading to labels ranging from 2 to 5, which were into a RF algorithm.

a.3) FAS Model 3: A hierarchical model consisting of two independent algorithms was developed. A first RF binary classifier to distinguish between lower (scores 1-2) and higher (scores 3-5) FAS scores, followed by a RF regression module to estimate a continuous value for the FAS scores between 3 and 5. Finally, the two hierarchical model parts were merged to assess the performance.

b.1) FAS Final Model: The individual task predictions of the best-performing model were aggregated using a linear model to obtain a total score as a percentage of the maximum achievable score across all tasks.

For the FAS Final Model, it was necessary to compute the maximum achievable scores for the Clinician True and Predicted Model 3 scores per task with Equation 3.6. The calculation takes into consideration that the maximum score that can be assessed per task trial is 5. Consequently, it was possible to compute the percentage of maximum achievable score for a patient at Baseline and Post-Treatment.

$$\%\, FAS\, max\, achievable\, score_{\,Task\, k} = \frac{\sum_{i=1}^{n} FAS_i}{5\, n} * 100 \tag{3.6}$$

where k is the Task, $k \in [0, 8]$; n is the total number of trials of the Baseline or Post-Treatment for a Subject in $Task\, k$; FAS_i is the i-th trial FAS score of the Baseline or Post-Treatment of $Task\, k$.

Afterward, all maximum achievable scores (in percentage) were aggregated considering the different tasks combinations as will be described in Table 4.10.

Regarding the FMA models, two models were evaluated to predict the single tasks trials scores. Furthermore, one model was employed to determine the final clinical scores prediction:

a.1) FMA Model 1: The FMA scores were grouped into classes for distance-based resampling purposes. The extracted features were utilized as inputs into a RF algorithm.

a.2) FMA Model 2: The FMA scores were grouped into classes for distance-based resampling purposes. The extracted features and the FAS predictions of the best-performing model were used as inputs into a RF algorithm.

b.1) FMA Final Model: The predicted FMA scores of the best-performing model were employed in a RF Regression algorithm which features inputs consisting of the eight tasks scores per subject Baseline and Post-Treatment assessments. The feature importance performed with the eight tasks enables the interpretation of the tasks combination that is clinically more suitable to estimate the FMA scores. Finally, a regression fit model was explored with the results from the previously mentioned algorithm.

For the FMA Final Model, since a subset of tasks translated to be more clinically suitable to estimate the Final FMA scores, that subset was used in the regression model fit. In the aforementioned, since there was solely one Clinician's True FMA score per patient Baseline and Post-Treatment visits, the patient trials average estimates for the Baseline and Post-Treatment visits were taken as input for the fit.

Chapter 4

Results

The following chapter presents the results accomplished in the analysis of an ABI dataset consisting of a total of thirty-seven patients (sixteen strokes and twenty-one traumatic brain injuries). The study investigated the two time points, Baseline, and Post-Treatment data assessments. The main findings regarding the participant's clinical data are described in Section 4.1. Furthermore, Section 4.2 stresses the pre-processing methodology outcomes. Finally, the FAS and FMA analysis of the estimated clinical scores focused on the upper-limb assessment are presented in respectively Sections 4.3 and 4.4. Notably, the models evaluated to predict the single and final clinical scores are examined.

4.1 ABI Participant's Clinical Data

The study encompasses a thirty-seven heterogeneous ABI participant's clinical dataset. Previously, were presented the clinical demographics of the sample (Table 3.1). Overall, the stroke and TBI group populations have the main difference in the age distribution. Figure 4.1 shows that TBI has a generally higher incidence in younger patients when compared to the stroke group, which frequently occurs in quinquagenarian to elderly patients.

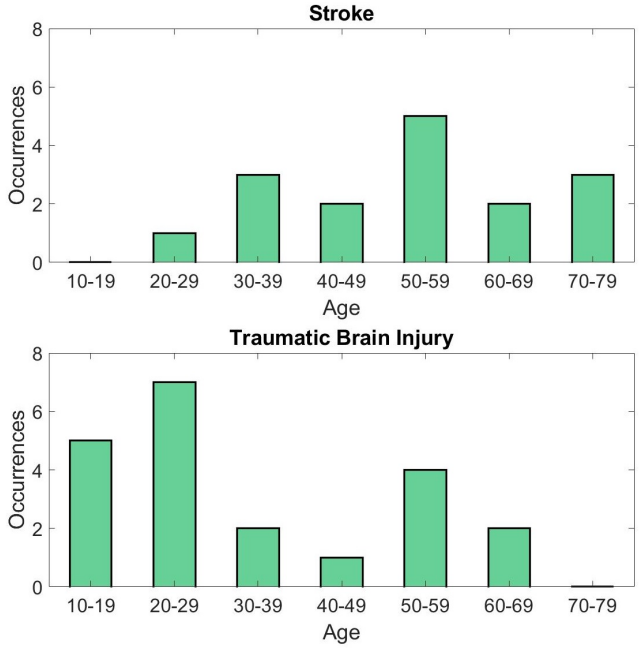


Figure 4.1: Stroke and TBI participant's age groups distribution profiles.

The mean FAS and FMA scores of the two ABI group populations are shown in Figure 4.2. What is interesting to note is that, for both clinical scores, on average, the Post-Treatment has higher scores. Namely, for the FAS scores, the stroke and TBI group at Post-Treatment revealed an increase of respectively 15% and 21% when compared to the Baseline, whereas for the FMA, the scores rose to 26% and 39%.

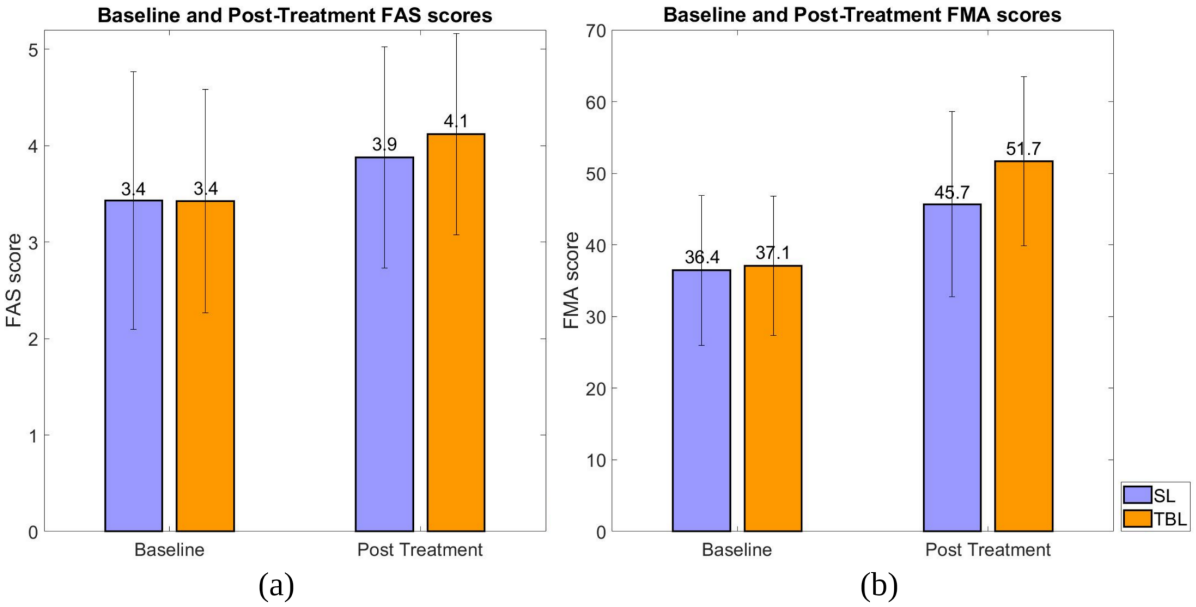


Figure 4.2: Mean (SD) clinical scores; on the left depicted the Baseline, and on the right, the Post-Treatment scores for both stroke lesion (SL) and traumatic brain lesion (TBL). The mean scores are presented in the bartop. (a) Mean FAS scores of the overall patient's task trials. (b) Mean FMA scores of the patients.

Consequently, when scrutinizing how the patients are recovering, there is not a noticeable difference between the stroke and TBI groups. The only contrast noted was the statistically significant age (i.e., the null hypothesis that the two groups are from populations with equal means is rejected) with a p-value$<$0.001 for a significance level of 0.05 in the statistical t-test.

Accordingly, for the purposes of analyzing this dataset, it was considered a patient-based analysis grouping stroke and TBI patients.

4.2 Signal Pre-Processing

To identify the axis of interest regarding the main wrist unit sensor located at the hemiparetic side and to whose index and thumb channels were connected, the channels' coefficients of correlation were computed. By analyzing data from different subjects and correlating the main wrist unit channels - Figure 4.3 illustrates this point clearly - it was possible to highlight that, generally, channel 6 was highly correlated to the wrist unit Z-axis. Accordingly, it was opted to consider only channels 7, 8, 9, and 10 as the fingers sensors' 4-axis containing the movements of interest.

Considering the calibration process, Figure 4.4 depicts the identification of the resting period window, given by the criteria mentioned in the methodology; in this case, the minimum RMS of the magnitude. Overall, this parameter enables detecting the window with the lowest signal variability, allowing the definition of the axis offsets in that window of interest.

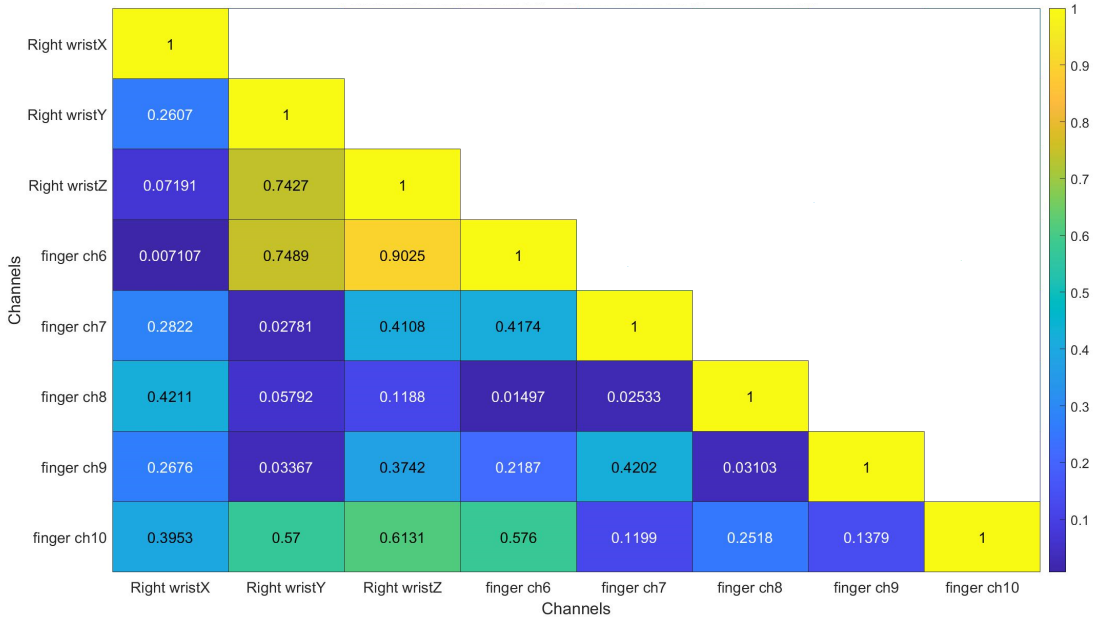


Figure 4.3: Matrix of correlation coefficients of the main wrist unit channels, including the finger's channels. The color scale represented on the left includes a soft fade from blue to yellow; where warm colors indicate higher correlation indices. The example depicted corresponds to Stroke patient 10, hemiparetic side wrist unit, in the Post-Treatment assessment at Task 1.

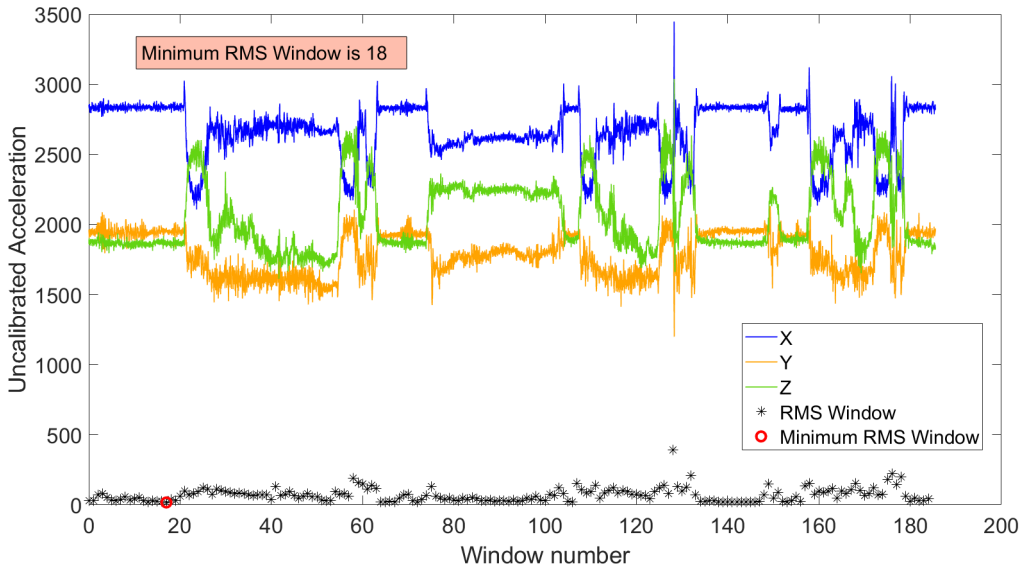


Figure 4.4: Minimum RMS window definition based on signal segmentation in windows. The example illustrated corresponds to Stroke patient 4, hemiparetic side upper-arm unit, in the Post-Treatment assessment at Task 1. The uncalibrated acceleration 3-axis is represented individually. The * corresponds to the windows (RMS of the magnitude). The red circle around window 18 indicates the minimum RMS window.

The calibrated data plots are represented in Figure 4.5, where it is clear that the magnitude of the acceleration is comparable to the expected magnitude acceleration at rest. The analysis was performed for all subjects where the mean value for each sensor resulted in $(9.81 \pm 0.11)\ m/s^2$ in the rest window. Altogether, this method can be considered a good calibration approximation.

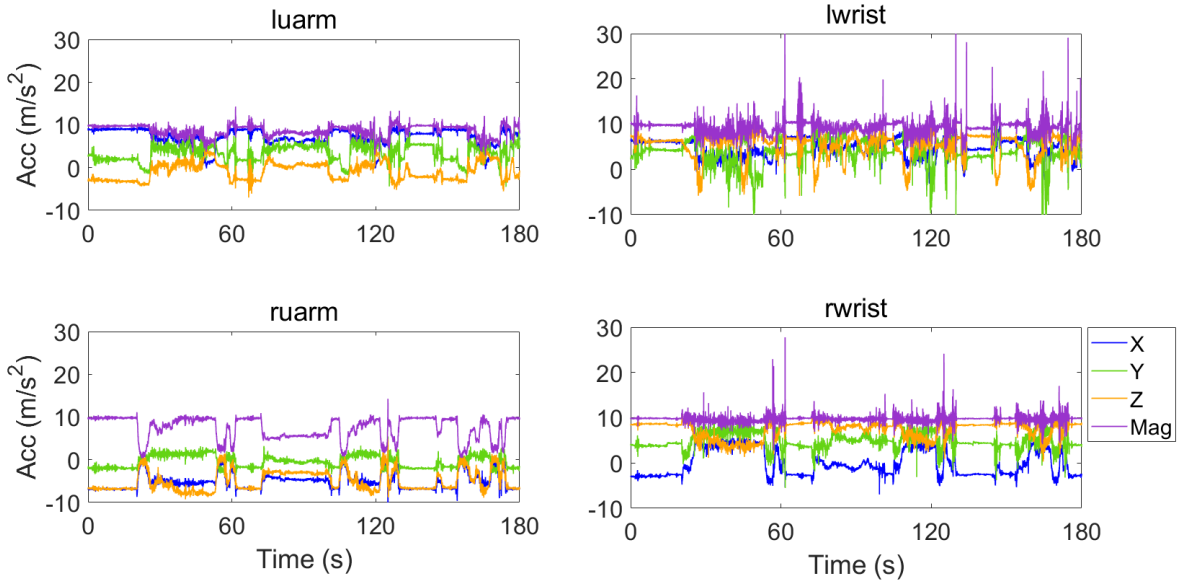


Figure 4.5: Calibrated acceleration data. The example illustrated corresponds to Stroke patient 4, upper-arm and wrist units, in the Post-Treatment assessment at Task 1. The calibrated acceleration 3-axis is represented individually. The purple line indicates the magnitude of the acceleration.

After calibration, a 6th-order Butterworth passband filter with 0.5 and 10 Hz cutoffs was applied to the dataset. These filtering applications aimed to narrow the frequencies of the signal to the ones that are clinically relevant while reducing the noise, distortion, and complexity of the signals, as can be clearly seen by Figure 4.6.

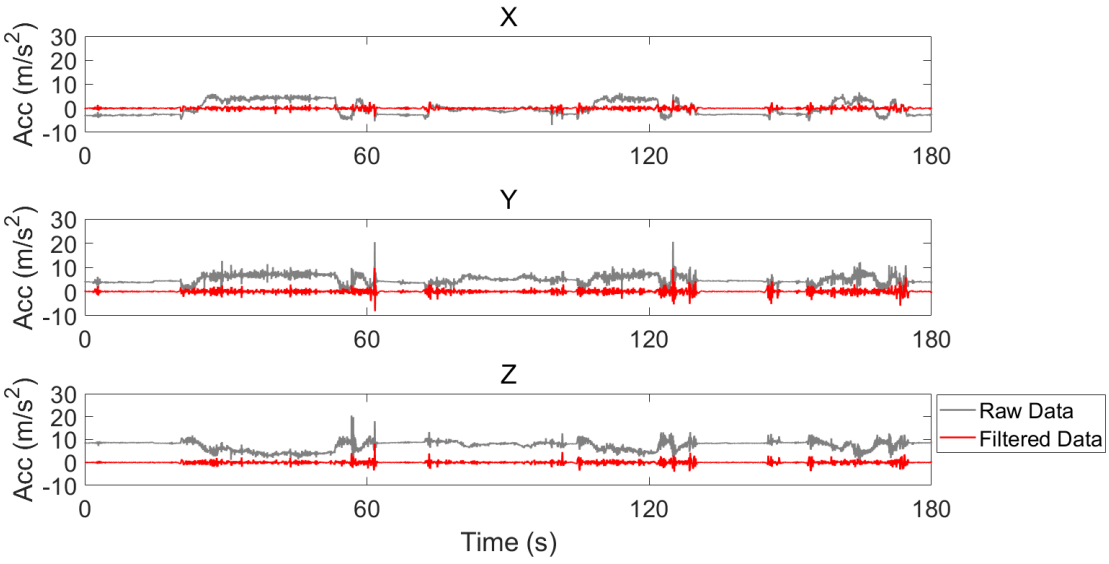


Figure 4.6: Filtered acceleration data with a 6th order Butterworth filter with passband between 0.5 and 10 Hz. The example illustrated corresponds to Stroke patient 4, hemiparetic wrist unit, in the Post-Treatment assessment at Task 1. The acceleration 3-axis is represented individually.

Moreover, by analyzing Figure 4.6, it is possible to identify the tree trials contained in each axis of the wrist sensor at Task 1. The full task time series went through trial segmentation using the markers. Accordingly, further analysis will be performed on a trial-by-trial basis.

4.3 FAS Estimation

The fundamental goal of this section is to present the results on estimating the FAS clinical scale aiming to assess the quality of movement more objectively using sensor technology. The clinician evaluated the FAS with six scores (0 to 5 points) per each trial of the task performance by the ABI patient.

4.3.1 Single Tasks Scores Prediction

The clinical distribution of the scores is imbalanced, as observed in Figure 4.7. The number of trials per task falls more around the higher scores, meaning that overall, patients are characterized by a higher motor functioning level. This result is in accordance with the previous results shown in Figure 4.2 where mean values for the FAS at the Baseline were observed at approximately 3 points and around 4 points for the Post-Treatment assessment. This aspect is crucial to understand; in general, it is clinically relevant to care for this type of motor evaluation in individuals who are relatively on the high motor function level; these individuals have the capability of generating good physiological patterns of motion and can broadly benefit more from the rehabilitation treatment.

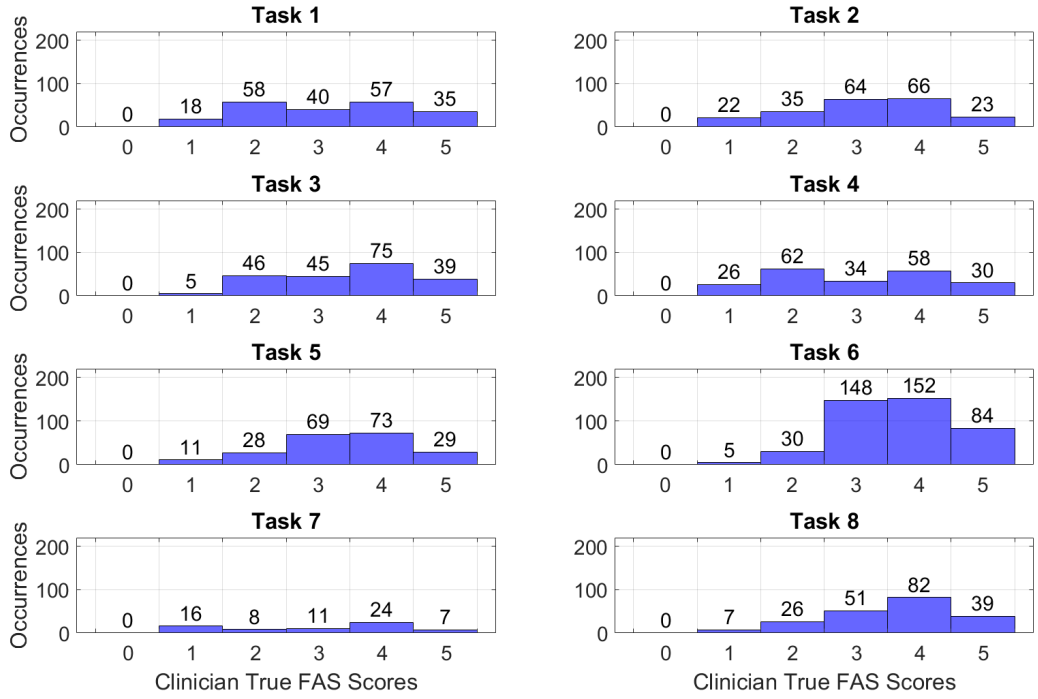


Figure 4.7: Clinician labeled FAS scores distribution per Task. The dataset includes the Baseline and Post-Treatment assessment of the individual trials. Note that each task comprised a different number of repetitions (see Figure 3.4).

Furthermore, in Figure 4.7, it is possible to note that across all tasks, no trials were assessed with 0 scores. Therefore, the analysis is reduced to a 5-point score problem. Likewise, the Figure reveals that some scores have few trials, particularly those labeled as 1 point have few occurrences for all tasks, excluding Task 7, where there was solely one repetition per subject assessment.

Accordingly, this imbalanced problem was faced using the ADASYN method; this method prevents the model from having a poor generalization and overfitting, which would be focused on the scores that have more observations.

Figure 4.8 exhibits the resampling results after ADASYN for Task 1, which are observed in the distribution of the various FAS clinical scores. This task represents an imbalanced dataset (see the superior panel of Figure 4.8); the mean distribution of observations for the real data is 42 ($\pm$ 15) occurrences.

In this case, scores 1, 3, and 5 represent the minority class examples. Consequently, for these classes, synthetic data points were generated. The resultant dataset is rebalanced (see the inferior panel of Figure 4.8); the observations distribution of each class is marked around an average of 57 ($\pm$ 2) observations. The consequential rebalanced dataset reduced the standard deviation by 13 points. Moreover, additional statistical analysis results indicate that for all features, the new dataset with resampled points samples population mean distribution is not statistically different from the original dataset samples population.

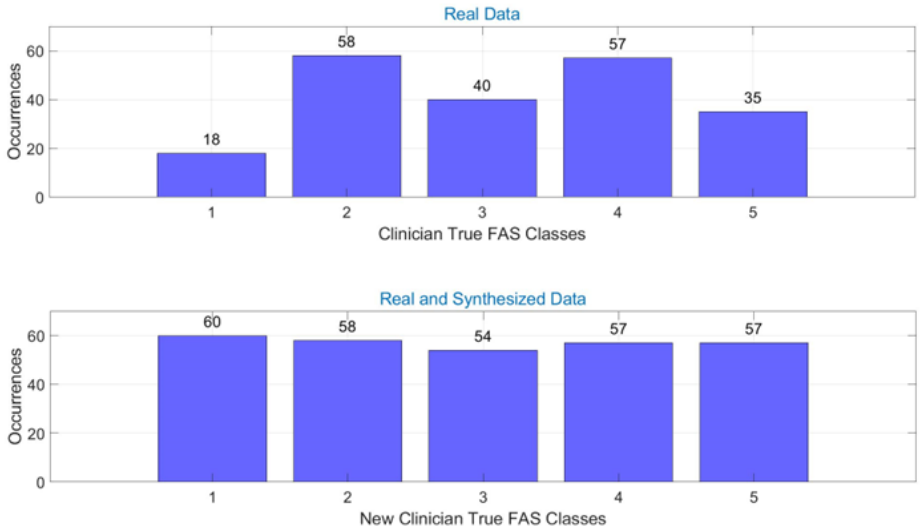


Figure 4.8: ADASYN synthetic point generation method applied for FAS rebalancing. The depicted example pertains to Task 1. The superior plot represents the imbalanced dataset consisting of the true clinician trial observations assessment of the FAS class. The inferior panel shows the rebalanced dataset with the ADASYN additionally synthesized points for classes 1, 3, and 5.

In order to understand how the newly resampled features are distributed in the dataset, it is helpful to visualize the feature maps using multidimensional data representations. The t-SNE method [73] was applied to the extracted features ($\approx$ 600), allowing the expression of this high-dimensional dataset, in this case, on a three-dimensional cloud (Figure 4.9).

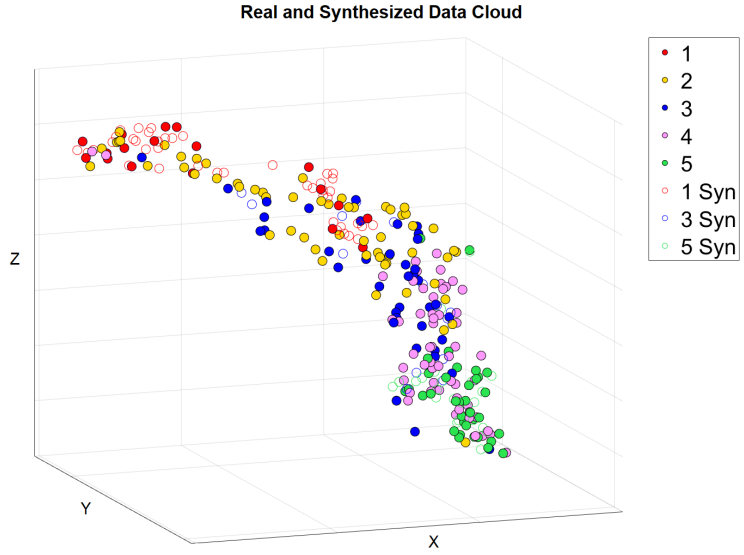


Figure 4.9: Data cloud representation using a t-Distributed Stochastic Neighbor Embedding, after ADASYN rebalancing. Real and synthesized data points are represented in a three-dimensional space and colored by the FAS score. The example pertains to Task 1.

As far as the data cloud is concerned, on the X axis, it is possible to observe that the points flow from lower scores to higher scores, trending from left to right. This is a fair indication that the features provide a hierarchical clustering of the FAS scores.

Figure 4.10 portrays a particular representation of the resampled points pertaining to the newly extracted features shown in Figure 4.9. What stands out is that the individual data clouds per class with the real and synthesized points represent a suitable generalization of the real data points cluster, in the sense that for all classes, the synthesized points blend agreeably with the real points distribution.

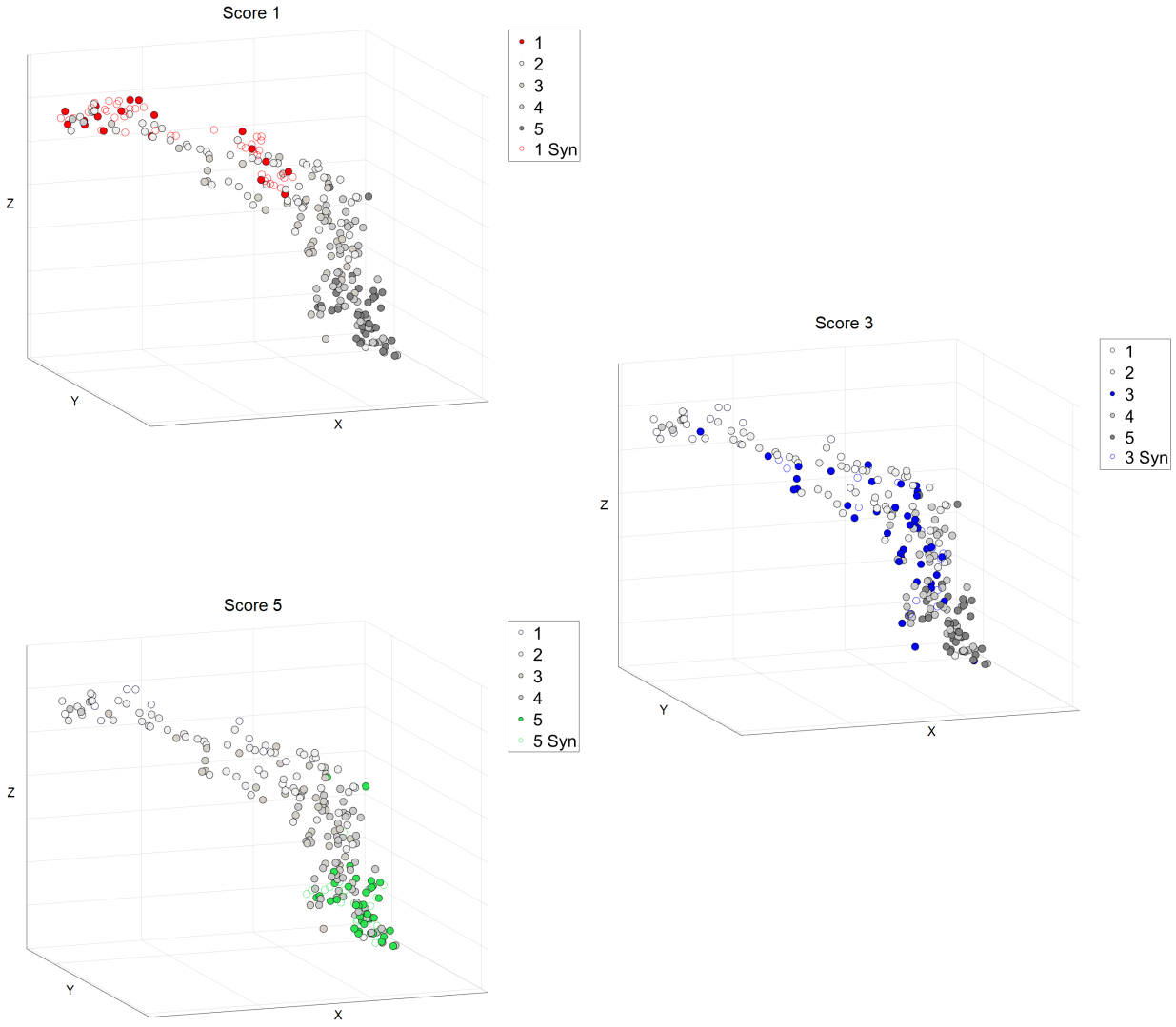


Figure 4.10: Individual FAS scores data clouds representation using a t-Distributed Stochastic Neighbor Embedding, after ADASYN rebalancing. Each cloud is colored by the FAS score and presents the real and synthesized points. The example pertains to Task 1.

What follows are the results of the analyses performed on the filter type and wrapper type supervised feature selection methods. The summary of the main findings is described in Table 4.1. To compare the models, the optimal number of training features, RMSE, accuracy, and F1 score metrics were calculated.

Broadly, for all motor Tasks, the wrapper type RF built-in feature importance method achieves more suitable performance with lower RMSE, higher accuracy, and F1 scores in comparison to the MRMR ReliefF and NCA filter type methods. Furthermore, the number of training features across tasks emerged to reduce the feature dimensionality and decrease computation time.

Table 4.1: Comparison of the various Feature Selection methods per Task on model performance. The model used a Regression approach to train the extracted features on FAS scores. The highlighted values represent the best model performance for each task.

Task #	Feature Selection Method	# of Training Features	RMSE	Accuracy	F1 Score
1	MRMR	66	0.76	0.51	0.46
	ReliefF	84	0.86	0.49	0.44
	NCA	93	0.89	0.46	0.41
	RF	54	**0.72**	**0.58**	**0.57**
2	MRMR	80	0.79	0.54	0.53
	ReliefF	69	0.85	0.53	0.46
	NCA	38	0.83	0.55	0.45
	RF	66	**0.77**	**0.60**	**0.55**
3	MRMR	77	0.92	0.55	0.42
	ReliefF	70	1.09	0.52	0.40
	NCA	66	0.91	0.46	0.36
	RF	60	**0.68**	**0.60**	**0.49**
4	MRMR	83	1.13	0.40	0.27
	ReliefF	78	1.17	0.41	0.31
	NCA	88	0.93	0.48	0.36
	RF	82	**0.88**	**0.51**	**0.42**
5	MRMR	83	0.64	0.59	0.40
	ReliefF	95	0.71	0.57	0.40
	NCA	79	0.68	0.61	0.44
	RF	81	**0.54**	**0.69**	**0.58**
6	MRMR	90	0.88	0.48	0.36
	ReliefF	84	0.85	0.51	0.37
	NCA	90	0.81	0.51	0.37
	RF	88	**0.66**	**0.57**	**0.42**
7	MRMR	37	1.03	0.32	0.26
	ReliefF	86	0.93	0.40	0.30
	NCA	67	0.96	0.38	0.28
	RF	67	**0.89**	**0.41**	**0.32**
8	MRMR	91	0.79	0.58	0.46
	ReliefF	74	0.92	0.51	0.40
	NCA	63	0.75	0.59	0.43
	RF	90	**0.70**	**0.64**	**0.49**

The feature selection performance for Model 3 step 2, focusing on higher-level patient scores estimates, is highlighted in Figure 4.11 denoting the Task 1 feature importance estimates. The RF built-in feature importance method was used to derive the results. It is interesting to note that typically, the top features for all tasks were most importantly related to the signal peaks number, energy around the dominant frequency, the correlations between the chest and the wrist sensors, smoothness, and the magnitude of the sensors. Regardless, wrist unit and upper-arm unit measures related appear to show the greatest predictive pertinence for single task predictions.

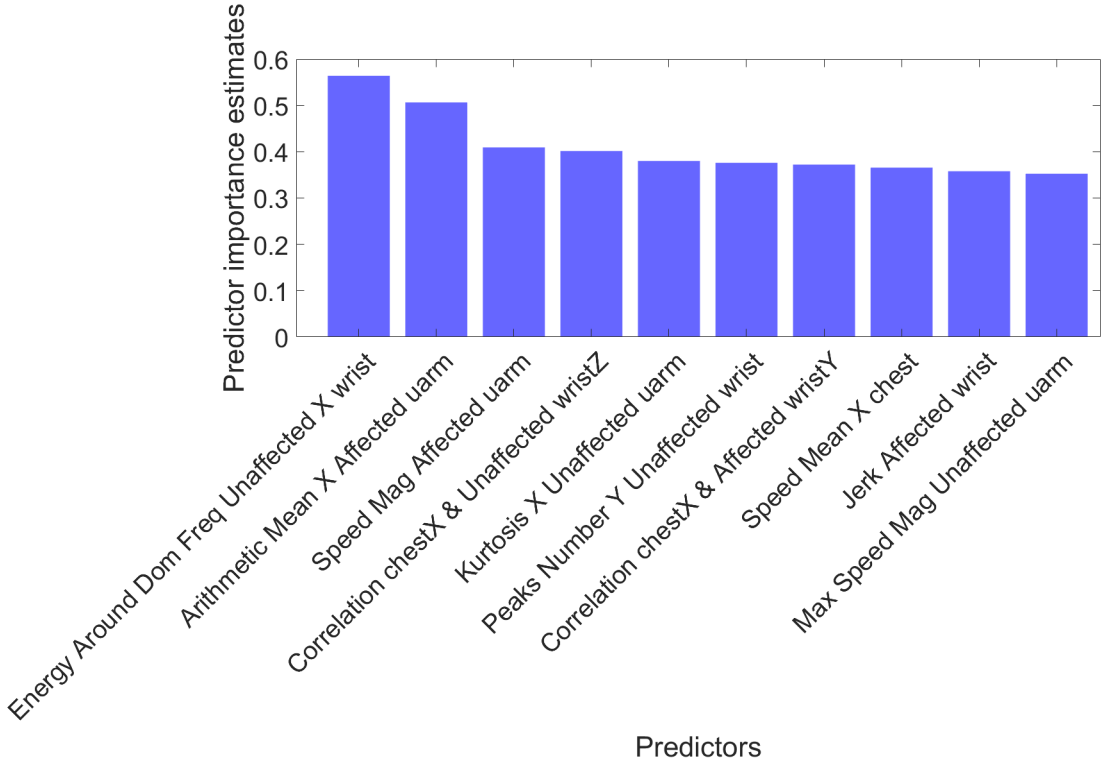


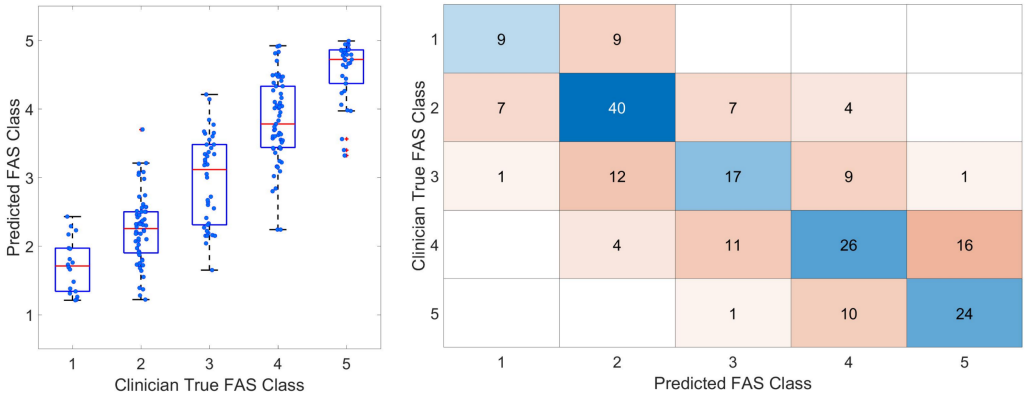
Figure 4.11: Feature importance for the top ten features for FAS Model 3 step 2 to predict FAS Task 1 Scores. The RF built-in feature importance and ADASYN rebalancing methods were employed to derive the resultant dataset.

The following represents the results for the single Tasks clinical scores estimates regarding the three models described in Subsection 3.2.5. The summary of the main findings regarding the models comparison can be analyzed and is highlighted in Tables 4.2 to 4.9. Whereas the particular case for the RF Regression and Classification pertaining to Model 1 is depicted in Figures 4.12, 4.14, 4.16, 4.18, 4.20, 4.22, 4.24, and 4.26.

Moreover, Model 3 emerged as optimal as a consequence of the metrics comparison with Model 1 and Model 2. The particular results for the Hierarchical model consisting of a primary RF Binary Classification followed by a RF Regression, are depicted in Figures 4.13, 4.15, 4.17, 4.19, 4.21, 4.23, 4.25, 4.27. Complementary results are presented in Appendix D.

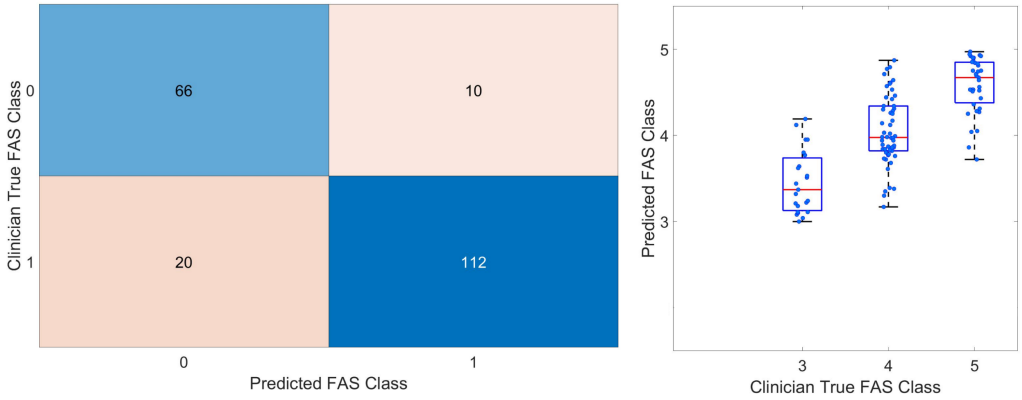
Table 4.2: Task 1 Models 1-3 to predict FAS Single Tasks Scores. Model performance is compared using a RF approach with RF built-in feature importance and ADASYN rebalancing. The highlighted values represent the optimal performance model's parameters.

Model	FAS Classes	Approach	RMSE	Accuracy	F1 Score
1	1,2,3,4,5	**Regression**	0.72	0.58	0.57
		Classification	0.63	0.56	0.55
2	2,3,4,5 (merged 1 and 2)	**Regression**	0.61	0.63	0.63
		Classification	0.65	0.63	0.61
3	Hierarchical, Step 1 1,2: labeled 0 3,4,5: labeled 1	**Binary Classification**	0.39	0.86	0.85
	Hierarchical, Step 2 3,4,5	**Regression**	0.39	0.86	0.85
		Classification	0.53	0.70	0.67
	Merged Hierarchical	Step 2 **Regression**	**0.31**	**0.68**	**0.64**



(a) (b)

Figure 4.12: Model 1 to predict Task 1 Single Task Scores concerning FAS classes 1 to 5. (a) RF Regression approach boxplots. Each Clinician True FAS class contains jitter points depicting the class trials. (b) Confusion Matrix concerning the RF Classification approach of the Clinician and Predicted FAS classes.

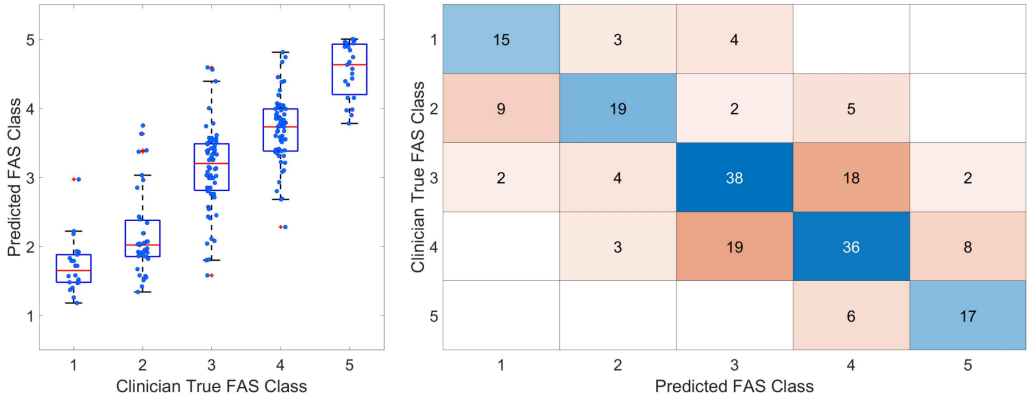


(a) (b)

Figure 4.13: Model 3 to predict Task 1 Single Task Scores concerning a Hierarchical group of FAS classes 1 to 5. (a) Step 1, Confusion Matrix concerning the RF Binary Classification approach of the Clinician and Predicted FAS classes (scores 1 and 2: labeled 0; scores 3 to 5: labeled 1). (b) Step 2, RF Regression approach boxplots. Each Clinician True FAS class contains jitter points depicting the trials of higher classes.

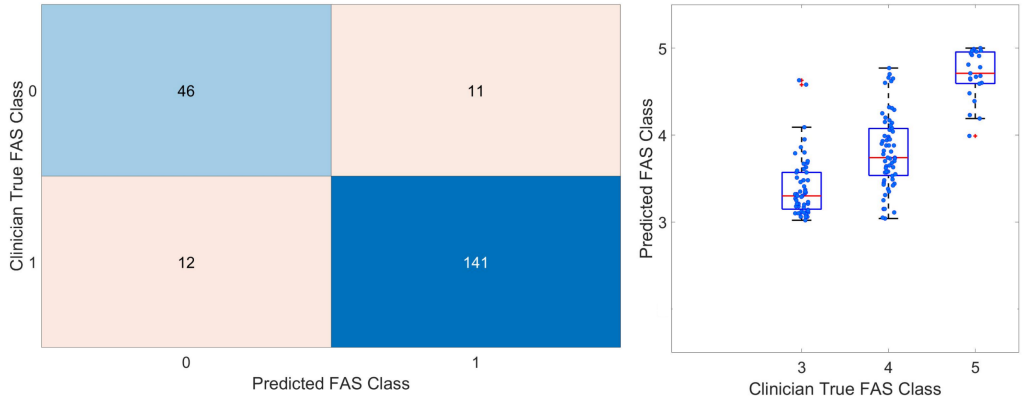
Table 4.3: Task 2 Models 1-3 to predict FAS Single Tasks Scores. Model performance is compared using a RF approach with RF built-in feature importance and ADASYN rebalancing. The highlighted values represent the optimal performance model's parameters.

Model	FAS Classes	Approach	RMSE	Accuracy	F1 Score
1	1,2,3,4,5	**Regression**	0.63	0.61	0.59
		Classification	0.51	0.60	0.60
2	2,3,4,5 (merged 1 and 2)	**Regression**	0.76	0.62	0.62
		Classification	0.74	0.67	0.64
3	Hierarchical, Step 1 1,2: labeled 0 3,4,5: labeled 1	**Binary Classification**	0.21	0.92	0.83
	Hierarchical, Step 2 3,4,5	**Regression**	0.28	0.89	0.86
		Classification	0.47	0.70	0.71
	Merged Hierarchical	Step 2 **Regression**	**0.32**	**0.68**	**0.69**



(a) (b)

Figure 4.14: Model 1 to predict Task 2 Single Task Scores concerning FAS classes 1 to 5. (a) RF Regression approach boxplots. Each Clinician True FAS class contains jitter points depicting the class trials. (b) Confusion Matrix concerning the RF Classification approach of the Clinician and Predicted FAS classes.

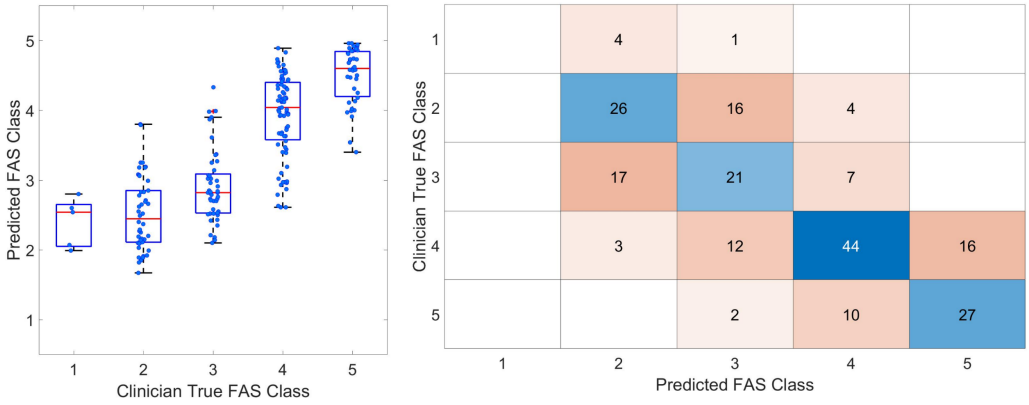


(a) (b)

Figure 4.15: Model 3 to predict Task 2 Single Task Scores concerning a Hierarchical group of FAS classes 1 to 5. (a) Step 1, Confusion Matrix concerning the RF Binary Classification approach of the Clinician and Predicted FAS classes (scores 1 and 2: labeled 0; scores 3 to 5: labeled 1). (b) Step 2, RF Regression approach boxplots. Each Clinician True FAS class contains jitter points depicting the trials of higher classes.

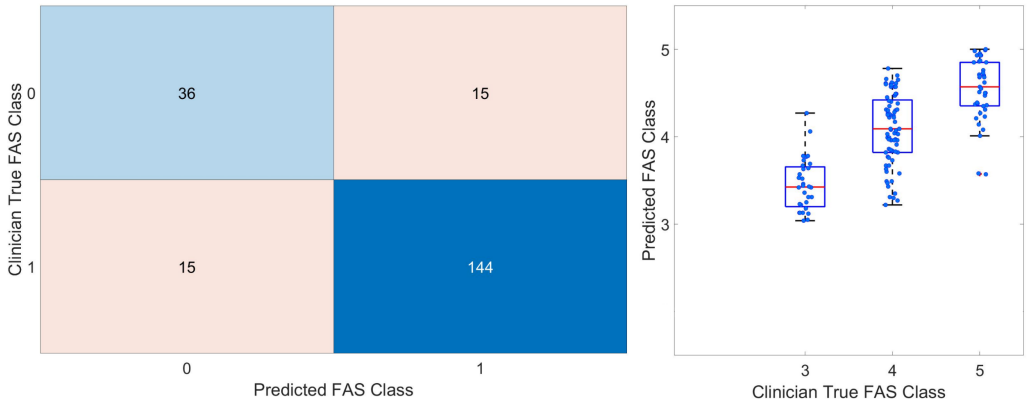
Table 4.4: Task 3 Models 1-3 to predict FAS Single Tasks Scores. Model performance is compared using a RF approach with RF built-in feature importance and ADASYN rebalancing. The highlighted values represent the optimal performance model's parameters.

Model	FAS Classes	Approach	RMSE	Accuracy	F1 Score
1	1,2,3,4,5	**Regression**	0.68	0.60	0.49
		Classification	0.68	0.56	0.45
2	2,3,4,5 (merged 1 and 2)	**Regression**	0.73	0.59	0.59
		Classification	0.64	0.60	0.60
3	Hierarchical, Step 1 1,2: labeled 0 3,4,5: labeled 1	**Binary Classification**	0.31	0.86	0.81
	Hierarchical, Step 2 3,4,5	**Regression**	0.31	0.86	0.81
		Classification	0.64	0.65	0.64
	Merged Hierarchical	Step 2 **Regression**	**0.39**	**0.61**	**0.60**



(a) (b)

Figure 4.16: Model 1 to predict Task 3 Single Task Scores concerning FAS classes 1 to 5. (a) RF Regression approach boxplots. Each Clinician True FAS class contains jitter points depicting the class trials. (b) Confusion Matrix concerning the RF Classification approach of the Clinician and Predicted FAS classes.



(a) (b)

Figure 4.17: Model 3 to predict Task 3 Single Task Scores concerning a Hierarchical group of FAS classes 1 to 5. (a) Step 1, Confusion Matrix concerning the RF Binary Classification approach of the Clinician and Predicted FAS classes (scores 1 and 2: labeled 0; scores 3 to 5: labeled 1). (b) Step 2, RF Regression approach boxplots. Each Clinician True FAS class contains jitter points depicting the trials of higher classes.

Table 4.5: Task 4 Models 1-3 to predict FAS Single Tasks Scores. Model performance is compared using a RF approach with RF built-in feature importance and ADASYN rebalancing. The highlighted values represent the optimal performance model's parameters.

Model	FAS Classes	Approach	RMSE	Accuracy	F1 Score
1	1,2,3,4,5	**Regression**	0.88	0.51	0.42
		Classification	0.70	0.50	0.47
2	2,3,4,5 (merged 1 and 2)	**Regression**	0.82	0.57	0.55
		Classification	0.75	0.61	0.57
3	Hierarchical, Step 1 1,2: labeled 0 3,4,5: labeled 1	**Binary Classification**	0.40	0.85	0.85
	Hierarchical, Step 2 3,4,5	**Regression**	0.40	0.85	0.85
		Classification	0.59	0.70	0.67
	Merged Hierarchical	Step 2 **Regression**	**0.30**	**0.68**	**0.61**

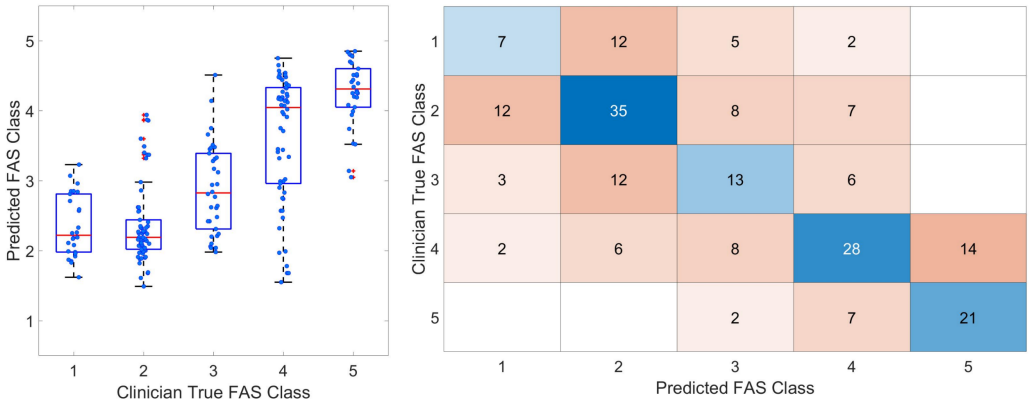


Figure 4.18: Model 1 to predict Task 4 Single Task Scores concerning FAS classes 1 to 5. (a) RF Regression approach boxplots. Each Clinician True FAS class contains jitter points depicting the class trials. (b) Confusion Matrix concerning the RF Classification approach of the Clinician and Predicted FAS classes.

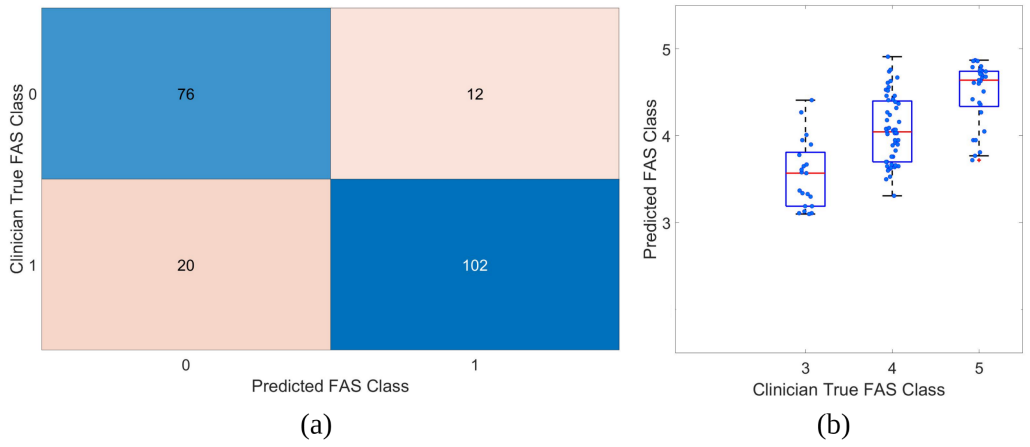
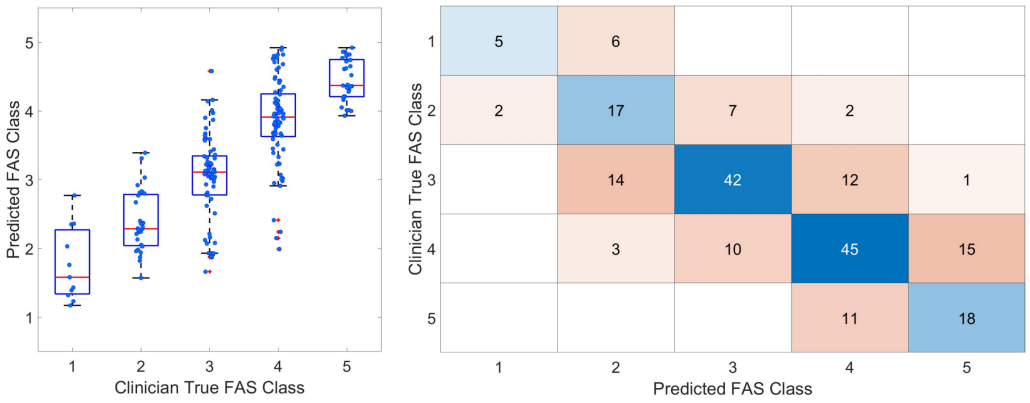


Figure 4.19: Model 3 to predict Task 4 Single Task Scores concerning a Hierarchical group of FAS classes 1 to 5. (a) Step 1, Confusion Matrix concerning the RF Binary Classification approach of the Clinician and Predicted FAS classes (scores 1 and 2: labeled 0; scores 3 to 5: labeled 1). (b) Step 2, RF Regression approach boxplots. Each Clinician True FAS class contains jitter points depicting the trials of higher classes.

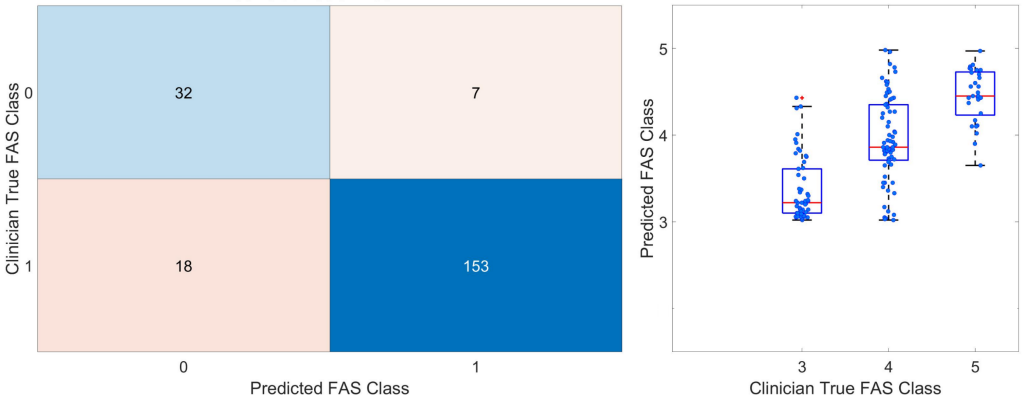
Table 4.6: Task 5 Models 1-3 to predict FAS Single Tasks Scores. Model performance is compared using a RF approach with RF built-in feature importance and ADASYN rebalancing. The highlighted values represent the optimal performance model's parameters.

Model	FAS Classes	Approach	RMSE	Accuracy	F1 Score
1	1,2,3,4,5	**Regression**	0.74	0.59	0.58
		Classification	0.62	0.60	0.58
2	2,3,4,5 (merged 1 and 2)	**Regression**	0.54	0.72	0.68
		Classification	0.51	0.73	0.70
3	Hierarchical, Step 1 1,2: labeled 0 3,4,5: labeled 1	**Binary Classification**	0.19	0.95	0.87
	Hierarchical, Step 2 3,4,5	**Regression**	0.32	0.88	0.82
		Classification	0.74	0.65	0.62
	Merged Hierarchical	Step 2 **Regression**	**0.38**	**0.62**	**0.60**



(a) (b)

Figure 4.20: Model 1 to predict Task 5 Single Task Scores concerning FAS classes 1 to 5. (a) RF Regression approach boxplots. Each Clinician True FAS class contains jitter points depicting the class trials. (b) Confusion Matrix concerning the RF Classification approach of the Clinician and Predicted FAS classes.

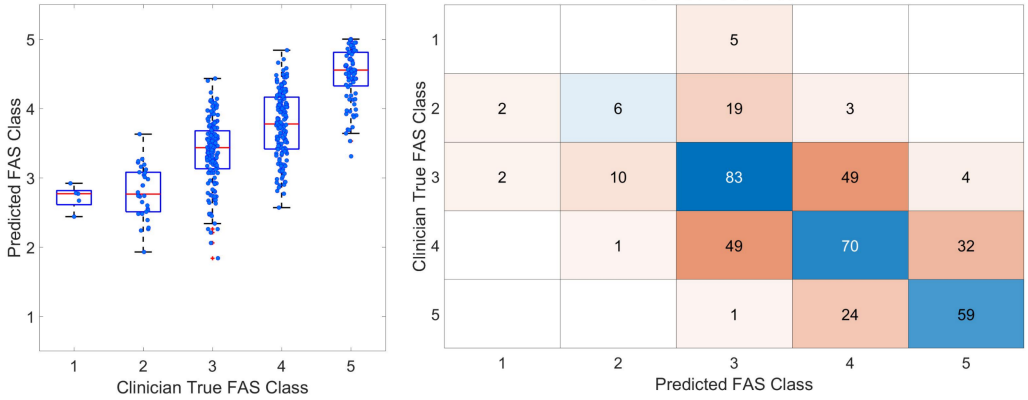


(a) (b)

Figure 4.21: Model 3 to predict Task 5 Single Task Scores concerning a Hierarchical group of FAS classes 1 to 5. (a) Step 1, Confusion Matrix concerning the RF Binary Classification approach of the Clinician and Predicted FAS classes (scores 1 and 2: labeled 0; scores 3 to 5: labeled 1). (b) Step 2, RF Regression approach boxplots. Each Clinician True FAS class contains jitter points depicting the trials of higher classes.

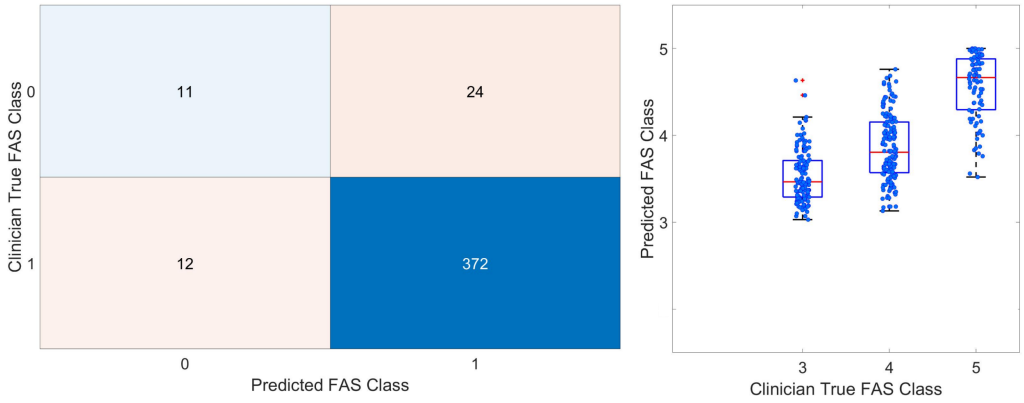
Table 4.7: Task 6 Models 1-3 to predict FAS Single Tasks Scores. Model performance is compared using a RF approach with RF built-in feature importance and ADASYN rebalancing. The highlighted values represent the optimal performance model's parameters.

Model	FAS Classes	Approach	RMSE	Accuracy	F1 Score
1	1,2,3,4,5	**Regression**	0.66	0.57	0.42
		Classification	0.58	0.52	0.39
2	2,3,4,5 (merged 1 and 2)	**Regression**	0.68	0.55	0.51
		Classification	0.68	0.52	0.51
3	Hierarchical, Step 1 1,2: labeled 0 3,4,5: labeled 1	**Binary Classification**	0.18	0.91	0.67
	Hierarchical, Step 2 3,4,5	**Regression**	0.18	0.91	0.67
		Classification	0.60	0.66	0.67
	Merged Hierarchical	Step 2 **Regression**	**0.39**	**0.61**	**0.58**



(a) (b)

Figure 4.22: Model 1 to predict Task 6 Single Task Scores concerning FAS classes 1 to 5. (a) RF Regression approach boxplots. Each Clinician True FAS class contains jitter points depicting the class trials. (b) Confusion Matrix concerning the RF Classification approach of the Clinician and Predicted FAS classes.

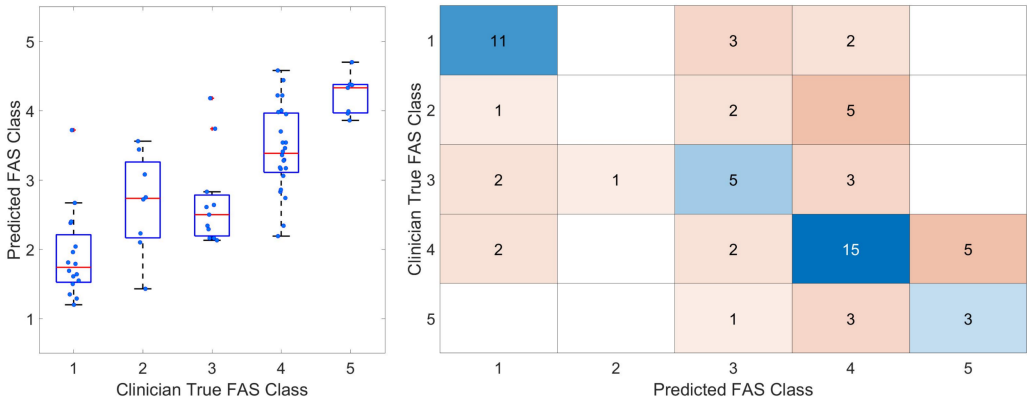


(a) (b)

Figure 4.23: Model 3 to predict Task 6 Single Task Scores concerning a Hierarchical group of FAS classes 1 to 5. (a) Step 1, Confusion Matrix concerning the RF Binary Classification approach of the Clinician and Predicted FAS classes (scores 1 and 2: labeled 0; scores 3 to 5: labeled 1). (b) Step 2, RF Regression approach boxplots. Each Clinician True FAS class contains jitter points depicting the trials of higher classes.

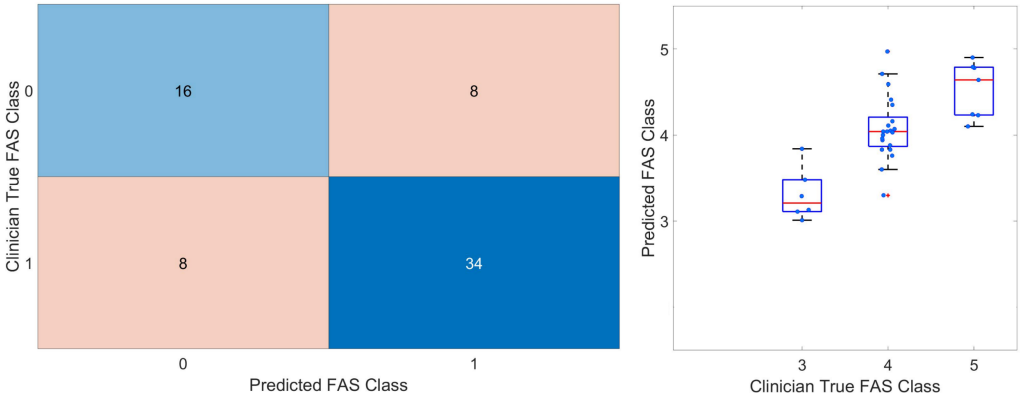
Table 4.8: Task 7 Models 1-3 to predict FAS Single Tasks Scores. Model performance is compared using a RF approach with RF built-in feature importance and ADASYN rebalancing. The highlighted values represent the optimal performance model's parameters.

Model	FAS Classes	Approach	RMSE	Accuracy	F1 Score
1	1,2,3,4,5	**Regression**	0.93	0.29	0.27
		Classification	0.93	0.52	0.52
2	2,3,4,5 (merged 1 and 2)	**Regression**	0.83	0.50	0.48
		Classification	0.58	0.65	0.58
3	Hierarchical, Step 1 1,2: labeled 0 3,4,5: labeled 1	**Binary Classification**	0.30	0.86	0.79
	Hierarchical, Step 2 3,4,5	**Regression**	0.44	0.76	0.74
		Classification	0.65	0.76	0.74
	Merged Hierarchical	Step 2 **Regression**	**0.36**	**0.61**	**0.58**



(a) (b)

Figure 4.24: Model 1 to predict Task 7 Single Task Scores concerning FAS classes 1 to 5. (a) RF Regression approach boxplots. Each Clinician True FAS class contains jitter points depicting the class trials. (b) Confusion Matrix concerning the RF Classification approach of the Clinician and Predicted FAS classes.

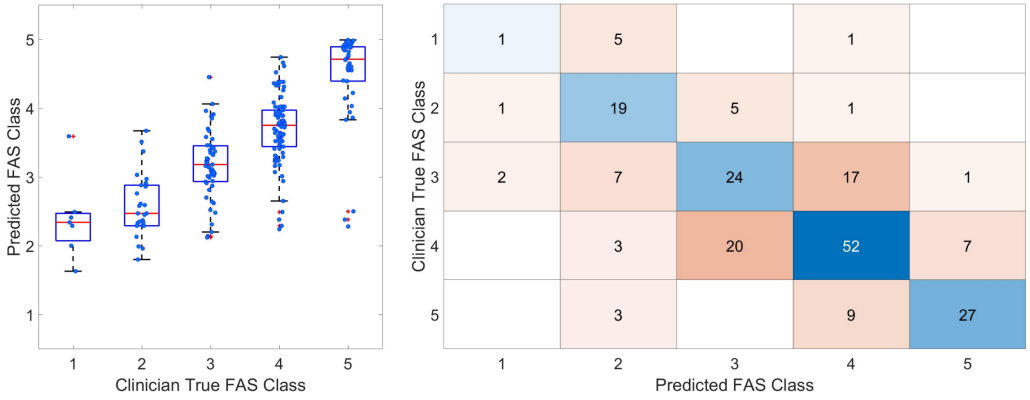


(a) (b)

Figure 4.25: Model 3 to predict Task 7 Single Task Scores concerning a Hierarchical group of FAS classes 1 to 5. (a) Step 1, Confusion Matrix concerning the RF Binary Classification approach of the Clinician and Predicted FAS classes (scores 1 and 2: labeled 0; scores 3 to 5: labeled 1). (b) Step 2, RF Regression approach boxplots. Each Clinician True FAS class contains jitter points depicting the trials of higher classes.

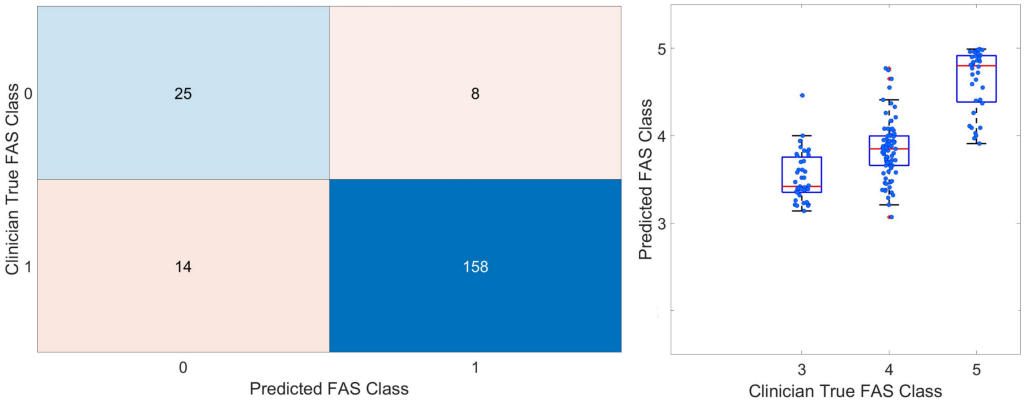
Table 4.9: Task 8 Models 1-3 to predict FAS Single Tasks Scores. Model performance is compared using a RF approach with RF built-in feature importance and ADASYN rebalancing. The highlighted values represent the optimal performance model's parameters.

Model	FAS Classes	Approach	RMSE	Accuracy	F1 Score
1	1,2,3,4,5	**Regression**	0.95	0.63	0.64
		Classification	0.96	0.60	0.53
2	2,3,4,5 (merged 1 and 2)	**Regression**	0.70	0.67	0.59
		Classification	0.58	0.71	0.63
3	Hierarchical, Step 1 1,2: labeled 0 3,4,5: labeled 1	**Binary Classification**	0.22	0.91	0.75
	Hierarchical, Step 2 3,4,5	**Regression**	0.29	0.89	0.81
		Classification	0.55	0.70	0.70
	Merged Hierarchical	Step 2 **Regression**	**0.34**	**0.64**	**0.64**



(a) (b)

Figure 4.26: Model 1 to predict Task 8 Single Task Scores concerning FAS classes 1 to 5. (a) RF Regression approach boxplots. Each Clinician True FAS class contains jitter points depicting the class trials. (b) Confusion Matrix concerning the RF Classification approach of the Clinician and Predicted FAS classes.



(a) (b)

Figure 4.27: Model 3 to predict Task 8 Single Task Scores concerning a Hierarchical group of FAS classes 1 to 5. (a) Step 1, Confusion Matrix concerning the RF Binary Classification approach of the Clinician and Predicted FAS classes (scores 1 and 2: labeled 0; scores 3 to 5: labeled 1). (b) Step 2, RF Regression approach boxplots. Each Clinician True FAS class contains jitter points depicting the trials of higher classes.

4.3.2 Combining Single Tasks Scores

The Hierarchical Model 3 was used to compute the final FAS estimation considering the individual tasks predictions.

A breakdown of the potential performance of a linear model in different combinations (i.e., subsets) of tasks was investigated. It is pertinent to note the Equation 3.6, described in Section 3.2.5, in order to interpret the results of the training of the linear regression model with all possible combinations of task scores. Across-the-board was considered the maximum achievable scores, in percentage, for the Clinician True and Predicted Model 3 scores per patient at Baseline and Post-Treatment.

The results for training a linear model with all possible combinations of tasks, ranging from a single task to all tasks, are presented in Table 4.10. The optimal fit considered for this model comprised all tasks, with a R^2 value of 0.91, as the outcomes indicate that all tasks are relevant for the optimal final prediction. Moreover, it was noted that, combined, some tasks conducted to better contribute to the final prediction, namely the bimanual Tasks 1 (Opening/Filling a Bottle), Task 3 (Putting on/Taking off Pen Cap), Task 6 (Lifting a box), and the unimanual Task 5 (Ironing).

Table 4.10: FAS aggregation considering Tasks combinations of the FAS scores of each patient at Baseline and Post-Treatment assessments. A fitted linear regression model was trained for the effect. The response and predictor variables considered were, respectively, the Clinician's True maximum achievable score and Predicted maximum achievable score (Model 3).

N.º of Combination	Optimal Tasks Combinations	R^2 max
1	5	0.62
2	3, 5	0.75
3	1, 3, 5	0.80
4	1, 3, 5, 6	0.84
5	1, 2, 3, 5, 6	0.87
6	1, 3, 4, 5, 6, 8	0.89
7	1, 2, 3, 4, 5, 6, 7	0.91
8	1, 2, 3, 4, 5, 6, 7, 8	0.91

Figure 4.28 depicts the latter, where the accomplished R^2 is 0.91 taking into account tasks from 1 to 8. It is noticeable that the adjusted fit line indicates that the predicted scores are close to the Clinician's True ones. Likewise, the lower and upper confidence bounds delimitate that some observations indeed fall in the interval, while globally, all observations fall closely around the confidence bounds.

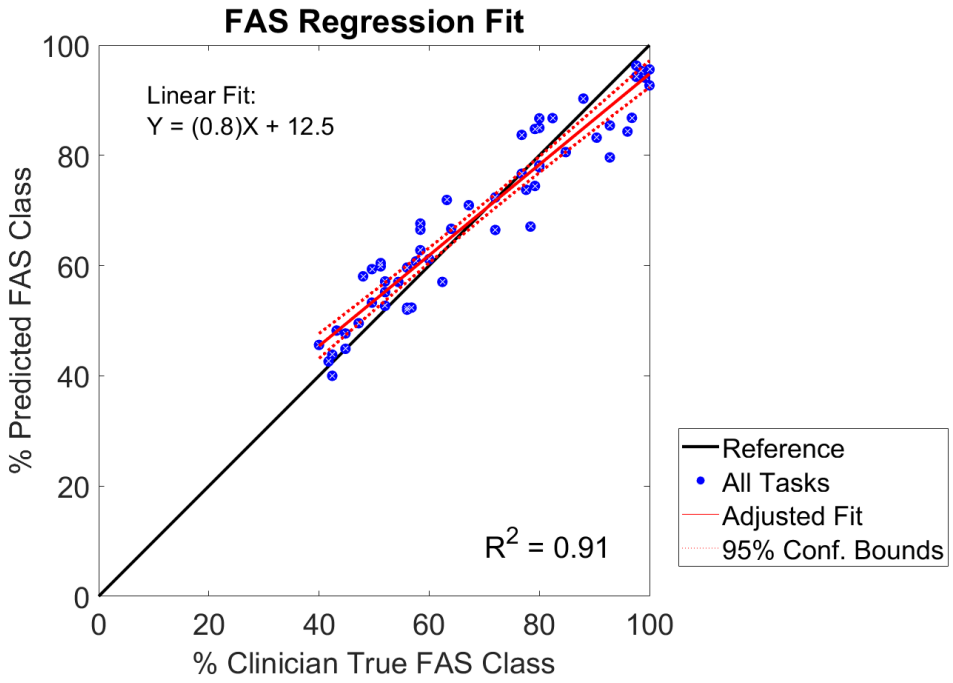


Figure 4.28: Scatter plot of the final estimate of the FAS scores of each patient at Baseline and Post-Treatment assessments. A fitted linear regression model was trained for the effect. The response and predictor variables considered were, respectively, the Clinician's True maximum achievable score and Predicted maximum achievable score. The final representation is relative to the percentage of the clinician labeled (true) FAS vs. the model 3 predicted FAS scores.

The particular results for the linear model derived in Figure 4.28 are presented in Table 4.11. The respective tasks coefficients (i.e., weights) reveal that, for some tasks, there is a stronger mathematical relationship between the FAS Clinician True and the dependent variable. Specifically, Tasks 1, 3, 5, and 6 have higher estimate values (0.13, 0.16, 0.17, 0.19, respectively), revealing higher weight importance in the final prediction. Considering the p-values, which provide information on the statistical significance of the relationship between the Clinical and Predicted scores, all tasks except for Task 8 have p-values of less than the 0.05 significance level. This provides sufficient evidence that there is a correlation between the changes in clinical labeled and predicted scores for Tasks 1 to 7, with a 95 % confidence interval.

Table 4.11: FAS linear regression model coefficients considering all Tasks of the FAS scores of each patient at Baseline and Post-Treatment assessments. The estimated correspond to the coefficient term; SE denotes the standard error of the coefficients; p-value provides information on the test null hypothesis that there is a zero correlation with the dependent variable (5% significance level).

Task Nº	Estimate	SE	p-value
1	0.13	0.03	< 0.001
2	0.10	0.04	0.03
3	0.16	0.03	< 0.001
4	0.11	0.03	0.002
5	0.17	0.05	0.001
6	0.19	0.04	< 0.001
7	0.07	0.03	0.02
8	0.06	0.04	0.12

The bias analysis of the above-presented data is shown in Figure 4.29 in order to evaluate the mean differences between the percentage of true and predicted scores. The resultant plot indicates a small bias and an estimation error in the interval of [-11,+12]% FAS points. Furthermore, most points fall into the confidence intervals ($\pm$ 1.96 SD). Likewise, is depicted the overestimation of lower scores and the underestimation of higher scores. The p-value achieved was 0.66, indicative that there is not sufficient evidence to identify a statistical significance in the estimation error (i.e., means difference between the scores).

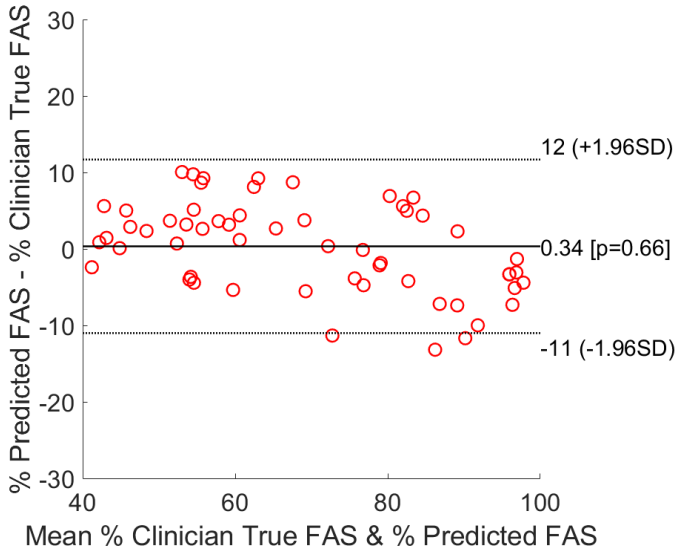


Figure 4.29: Bland-Altman plot the percentage of FAS Clinician's True and predicted scores, evaluating the bias between the patient's mean differences. The difference between the percentage of Predicted and Clinician True scores presents the estimation error. The sign of the estimation error denotes the under and overestimation.

4.4 FMA Estimation

The FMA estimation results were obtained with a dataset of FMA scores (0 to 66 points) assessed by the clinicians for each ABI patient at Baseline and Post-Treatment time points. In addition, the previously wearable accelerometer extracted features were employed, and additional FAS scores from Model 3 were investigated as features.

The correlation between the Estimated FAS scores and the Clinician's True FMA scores was computed as a benchmark (i.e., point of reference). The resultant coefficient of determination was 0.79, indicating a positive association between the variables. Hence, the aforementioned suggests that given an increase in FAS, there will likewise be a corresponding increase in the FMA score, consequently validating a direct relationship between the clinically labeled scores (see Figure 4.30).

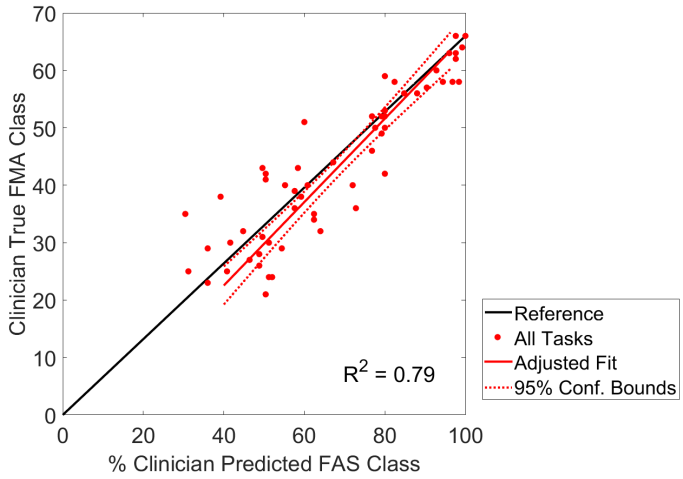


Figure 4.30: Scatter plot of the Clinician Predicted FAS scores and the FMA Clinician True scores of each patient at Baseline and Post-Treatment assessments. A fitted linear regression model was trained for the effect. The response and predictor variables considered were, respectively, the % Clinician Predicted FAS scores and FMA Clinician True scores.

4.4.1 Single Tasks Scores Prediction

These results' clinical distribution of the scores of the FMA are similar to those on the FAS distribution, where there is notably an imbalanced dataset (see Figure 4.31). As mentioned previously, the FMA is assessed on specific criteria concerning the proximal and distal parts of the upper-limb. The total FMA is the sum of the two proximal and distal components. As depicted in Figure 4.31, there is a noticeable range of lower FMA scores that present no observations. The minimum total score assessed was 21 points, and the maximum corresponds to the highest score of 66 points. The total mean of the scores falls close to 44 points, indicating that the occurrences are typically in the mid-range of the FAS motor impairment. The results support the observed in Figure 4.2 for the Baseline, and Post-Treatment mean FMA scores.

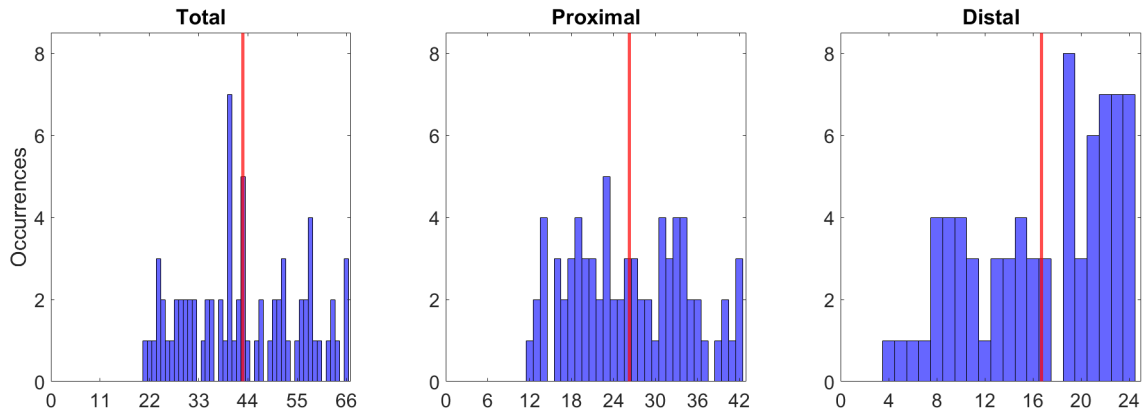


Figure 4.31: Clinician labeled FMA scores distribution comprising one score per subject Baseline and Post-Treatment assessment. The red line denotes the mean of the dataset. The proximal and distal scores, which sum to the total score, are individually depicted.

Similarly, as what succeeded with the FAS scores, for the FMA was necessary to apply a synthetic point generation method to balance the dataset. Particularly, for this clinical score was advantageous for grouping the points into classes. For the effect, five classes were chosen (Table 4.12). The class representation is shown in Figure 4.37 where it is explicit that the grouped scores have a reasonable number of occurrences for all class distributions.

Table 4.12: FMA five classes ranges for ADASYN rebalancing on a RF algorithm.

FMA Classes	Range
1	FMA ≤ 30
2	$30 <$ FMA ≤ 38
3	$38 <$ FMA ≤ 47
4	$47 <$ FMA ≤ 56
5	FMA > 56

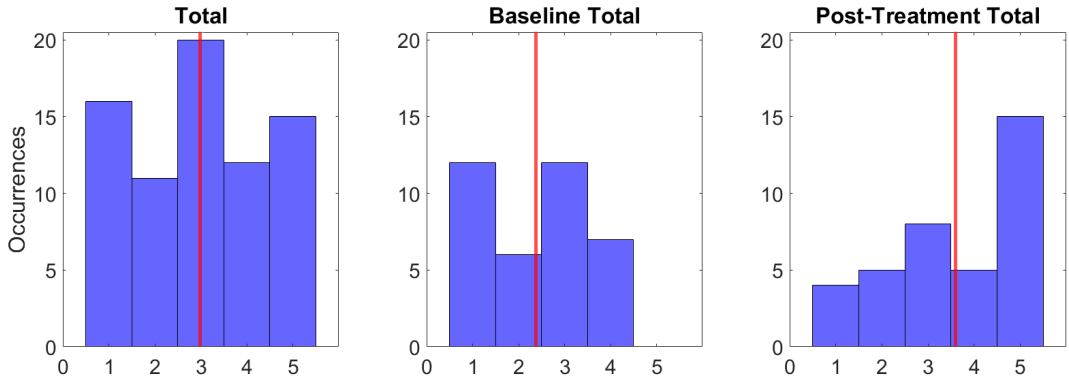


Figure 4.32: Clinician labeled FMA scores distribution in five classes comprising one score per subject Baseline and Post-Treatment assessment. The red line denotes the mean of the dataset. The Baseline and Post-Treatment scores, which sum to the total score, are individually depicted.

Figure 4.33 demonstrates the results for the employed ADASYN method in Task 1 FMA scores distribution per grouped class. In this example, scores 1,2,4, and 5 represent the minority class examples. The real data points in this task include a mean distribution of observations of 42 (± 8) occurrences; after resampling, the results are 53 (± 5) occurrences. Therefore, for this case, the real dataset does not suggest a high imbalance as the rebalancing process reduced the standard deviation by 3 points.

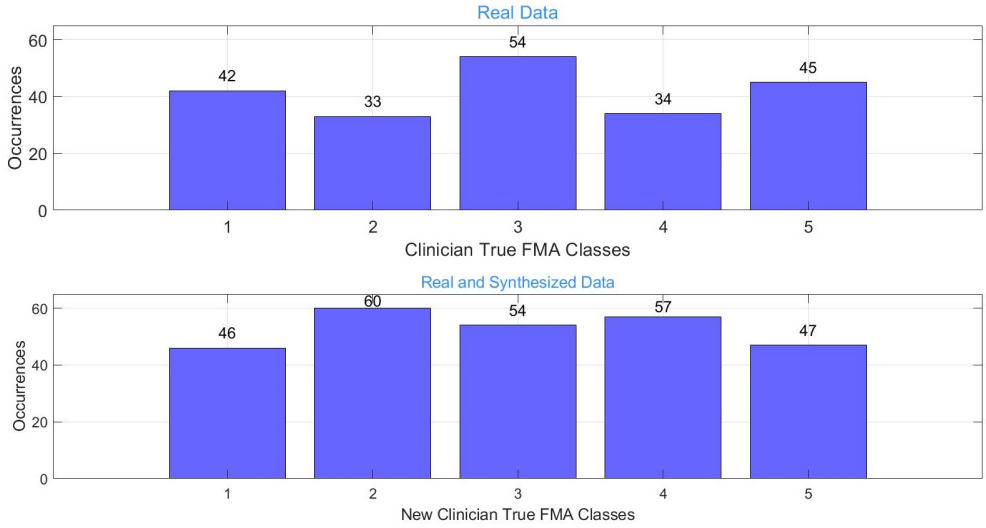


Figure 4.33: ADASYN synthetic point generation method applied for FMA rebalancing. The FMA scores were grouped into 5 classes. The depicted example pertains to Task 1, where FMA was labeled per task trial. The superior plot represents the imbalanced dataset consisting of the true clinician trial observations assessment of the FAS class. The inferior panel shows the rebalanced dataset with the ADASYN additionally synthesized points for classes 1, 2, 3, and 5.

For interpreting the estimated single FMA scores per patient Baseline and Post-Treatment, FMA models 1 and 2 were analyzed. The main contrast between the two models is that Model 1 doesn't consider the FAS as a feature to predict the estimates (i.e., the FAS is not taken into account as affecting the final results). The performance of Model 1 per task and Subject (see Appendix E) was poorer when compared to Model 2. Thus, the focus of the analysis is performed on Model 2.

The feature selection performance for Model 2, highlighted in Figure 4.34, denotes the Task 1 feature importance estimates. It is remarked that the predicted FAS score feature from the hierarchical Model 3 is the most important feature among all features computed. The propensity for the FAS scores feature being at the top, specifically in the top three predictor importance, was noted for all eight tasks (Table 4.13). Moreover, it was noted that typically, the top features were mainly related to the magnitude of the sensors on the upper-arm. The energy around the dominant frequency, associated with the periodicity of the movement, is presented as of high importance. In the depicted case, the chest Z axis also reflects importance. Nevertheless, wrist unit measures related to the magnitude were likewise of great predictive pertinence for single task predictions.

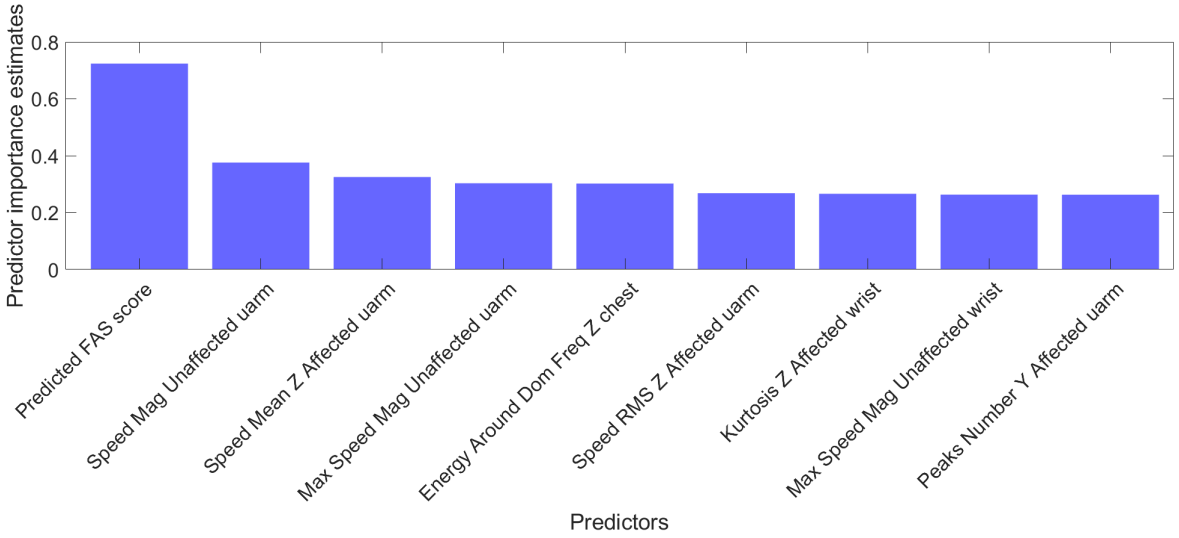


Figure 4.34: Feature importance for the top nine features for FMA Model 2 to predict FMA Task 1 Scores. The RF built-in feature importance and ADASYN rebalancing methods were employed to derive the resultant plot.

Table 4.13: Predicted FAS score feature (from FAS Model 3) estimated importance per Task assessed for FMA Model 2 with the Random Forest built-in feature importance method.

Task Nº	FAS Feature Ranking
1	1
2	1
3	3
4	1
5	3
6	1
7	1
8	1

The results for deriving FMA scores using the knowledge provided by the FAS estimates in Model 3 are outlined in Figures 4.35 and 4.36. What stands out in the analysis of Figure 4.35 is the general pattern of the RMSE per task. For instance, the lowest RMSE is 5.9 and verified for Task 1, the disparity to the highest RMSE verified in Task 8, which is approximately 1 point. Figure 4.36 depicts a similar analysis performed on a subject-by-subject basis where the variability of the parameter is marked.

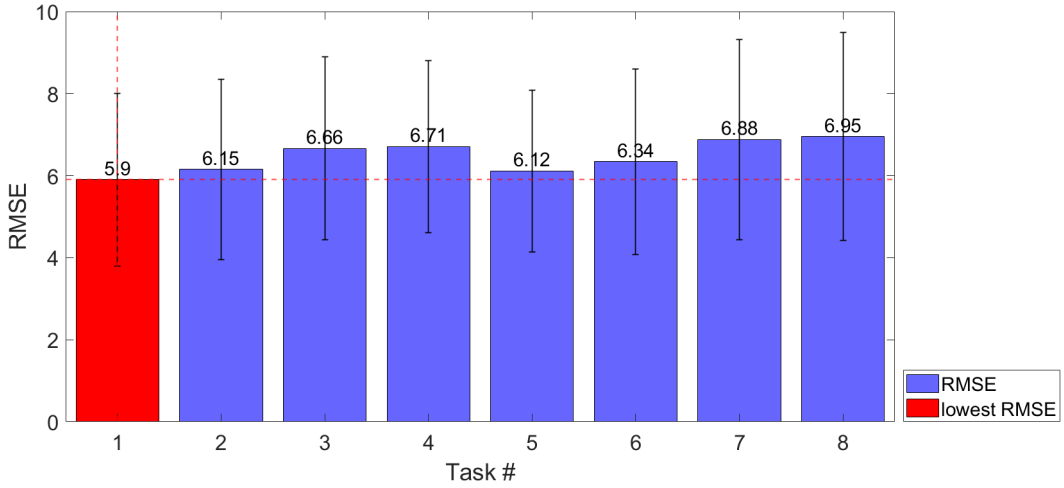


Figure 4.35: RMSE of the FMA Single Tasks Scores estimated with FMA Model 2. Emphasized in red is the lowest RMSE taking into account all Tasks comparisons.

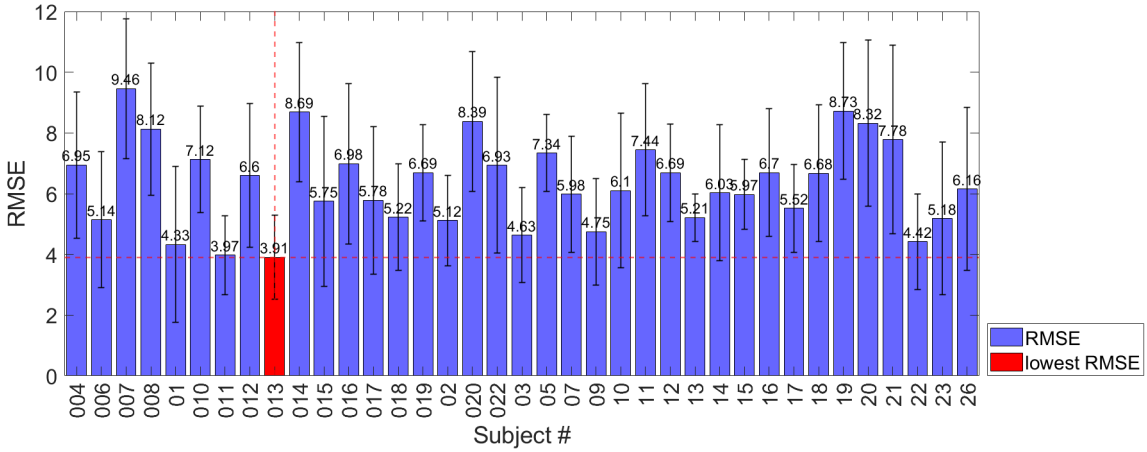


Figure 4.36: RMSE of the FMA Single Tasks Scores estimated with FMA Model 2 for each patient. Emphasized in red is the lowest RMSE taking into account all subject comparisons.

4.4.2 Combining Single Tasks Scores

The FMA Model 2 estimates were used to compute the final FMA estimation considering the individual Tasks predictions.

The predicted FMA scores of Model 2 were employed in a RF Regression algorithm which features inputs consisting of the eight Tasks scores per subject Baseline and Post-Treatment assessments. Figure 4.37 illustrates the optimal Tasks selected as translating to be the more suitable to explain the tracked movements regarding the severity of motor impairments assessed with FMA.

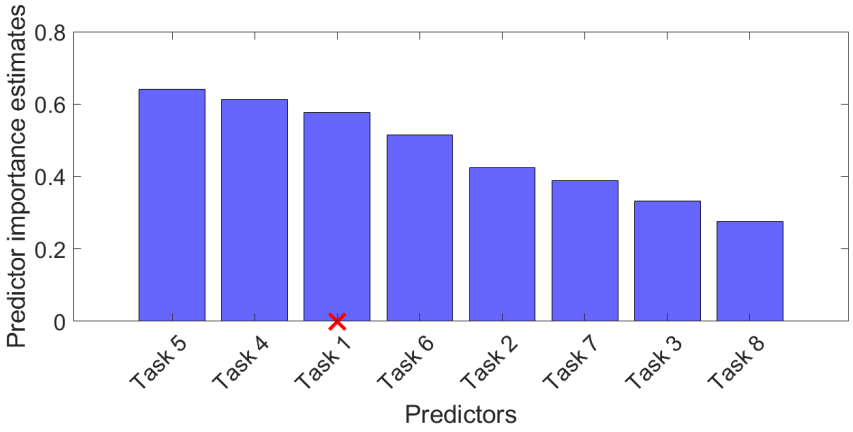


Figure 4.37: Feature importance regarding the RF model for the final estimate of the FMA scores of each patient at Baseline and Post-Treatment assessments. The Random Forest built-in feature importance method selects the optimal combination of task features marked in red.

A linear fit model was explored considering the results from the previously mentioned model. Figure 4.38 shows the adjusted fit with a coefficient of determination, R^2, of 0.79, resulting in the smallest difference between the clinically observed FMA scores and their fitted values.

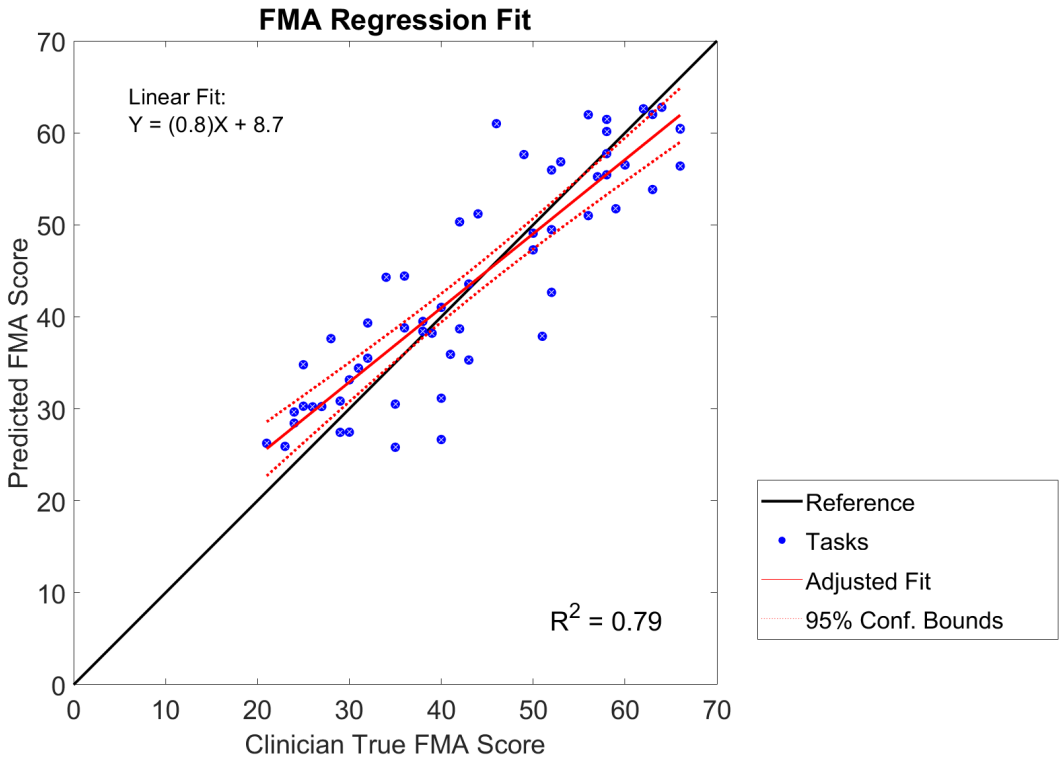


Figure 4.38: Scatter plot of the final estimate of the FMA scores regarding the optimal combination of tasks (i.e., tasks [1 4 5]) of each patient at Baseline and Post-Treatment assessments. The representation is relative to the clinician labeled (true) FAS vs. the model 3 predicted FAS scores.

Nonetheless, the above-presented regression line consistently over-predicts the lower scores and under-predicts the higher scores along the fit. Thus, it is desirable to conduct a bias analysis. Figure 4.39 reveals a Bland-Altman graph to evaluate the mean differences between the true and predicted scores. The results portray that the estimation error falls around the interval of $\pm$ 12 FMA points, where most points fall in the confidence intervals of ($\pm$ 1.96 SD). Consequently, there is an increase in the

number of observations in lower scores, suggesting an overestimation error; for higher scores, a slight underestimation is noted. The p-value achieved was 0.74, indicative of a small bias and informative that there is insufficient evidence to identify a statistical significance in the estimation error (i.e., means difference between the scores) at a 96 % level of confidence.

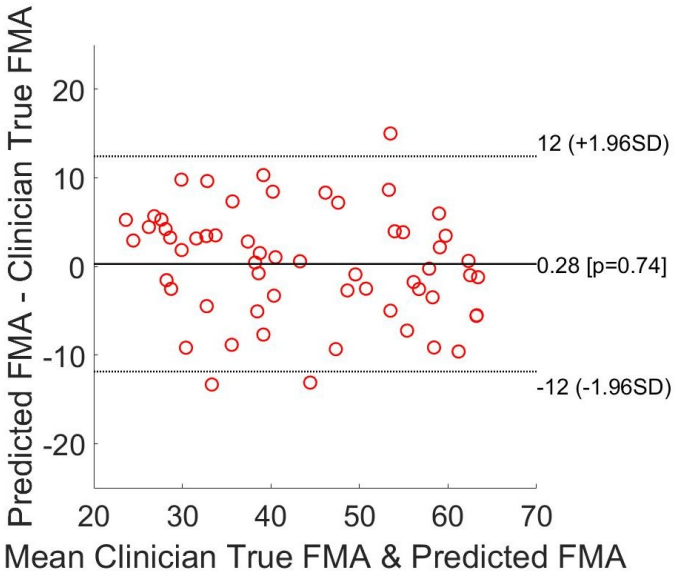


Figure 4.39: Bland-Altman plot the FMA Clinician's True and predicted scores, evaluating the bias between the patient's mean differences. The difference between the Predicted and Clinician True scores presents the estimation error. The sign of the estimation error denotes the under and overestimation.

To better scrutinize the resultant FMA bias, the FMA scores estimation error was analyzed as a function of the Clinician's True FMA as observed in Figure 4.40. In addition, the Minimal Detectable Change (MDC) estimation error (i.e., the smallest detectable difference calculated by the authors and which might be considered to be a true change rather than a measurement error) was used as a reference as defined in the literature [10].

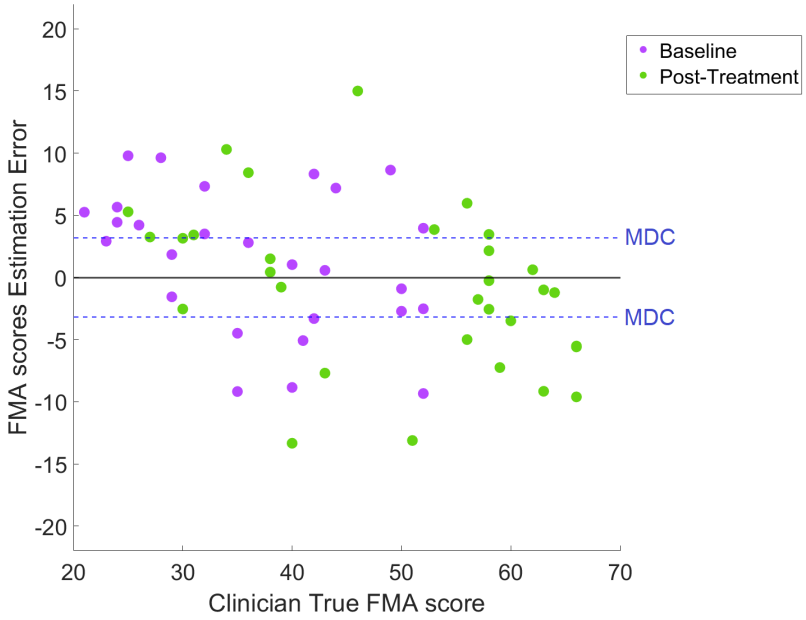


Figure 4.40: Error estimation of the FMA Clinician's True and predicted scores colored by the Baseline or Post-Treatment assessments. The FMA Minimal Detectable Change (MDC) reference boundaries are defined [10].

Chapter 5

Discussion

The increasing global incidence of stroke and TBI neurological injuries represents a public health issue with great contributor to persistent disability amongst the survivors [30], [21]. Moreover, upper-limb hemiparesis is a frequent sequelae of these events, limiting the performance of daily activities of the survivors [26],[27]. Consequently, delivering suitable rehabilitation plans focused on upper-limb programs is crucial to promoting motor gains that empower patients to regain confidence in ADL performance. This work aims to develop new automation techniques to assess ABI clinically relevant motor outcomes. For this purpose, wearable sensors tracked and quantified patients' upper-limb movements while performing ADLs, enabling a non-invasive and easy assessment of the patients in a clinical setting. Furthermore, several algorithms were developed to optimally combine the sensing technology data in order to derive the final clinical relevant outcomes, namely FAS and FMA scores. Ultimately, this study demonstrates the feasibility of providing the healthcare professional with novel tools to assess upper-limb rehabilitation outcomes more consistently over the care program and tailor the patient-specific plans accordingly.

5.1 Clinical FAS and FMA Automation for ABI Assessment

Consistently with the previous literature, Figure 4.1 denotes that the results align with the review, where stroke incidence increases with age-related comorbidities, whereas TBI has primary risk factors for persons under twenty-years-old, potentially associated with automobile, and sports accidents [22]. Furthermore, the observed high incidence of TBI in quinquagenarians is potentially associated with transport accidents, while elderly patients show a high level of falls [18].

The mean FAS and FMA scores of the two ABI group populations are shown in Figure 4.2. What is interesting to note is that, for both clinical scores, on average, the Post-Treatment has higher scores. Hence, this is a fair indication of the success of the upper-limb treatment, which led to improvements in the clinical performance of the second data acquisition recording. Post-Treatment revealed an increase of 15% and 21% points compared to the Baseline; whereas for the FMA, this increase was 26% and 39%. The increased improvement in Post-Treatment assessment for TBI patients might result from the statistically significant age difference between the subjects of the two pathologies. Broadly, younger TBI patients show higher compensation capacity and higher cognitive reserves [74].

5.1.1 FAS Estimation

The FAS assesses the patient's body structure and functioning as well as activity in the ICF frame-

work. Automating this clinical scale using instrumental devices on the ABI population, specifically focusing on the upper-limbs, enables clinicians to analyze the patient's quality of movement over the plan of care.

The ABI participants of this study were assessed with the FAS clinical scale at the traditional two time points: Baseline and Post-Treatment sessions. The assessment at these two time points is of particular relevance because, after a rehabilitation intervention program focused on the upper-limb, an improvement in the quality of movement is expected. In addition, the laboratory was equipped with a clinical setting that simulated the particular tasks. Assessing the FAS at the eight performing tasks and all trials of the specific task enables capturing the particular movement of interest, reducing the variability of the results.

The wearable devices used to acquire the FAS data enabled patients to perform activities unobtrusively due to the particular specifications of the Shimmer2r sensors, including being compact and lightweight.

Taking these considerations into account, the algorithm to estimate the clinical scale initially focused on estimating each of the eight task trials' FAS scores. As explained previously in Section 3, a supervised ML RF-based algorithm where the clinical labels could be introduced was preferred. For the scope of this Dissertation, after analyzing a different number of trees, 100 decision trees are evaluated as the optimal since as the number of trees increased, no significant improvements in the models were reached while fewer computational costs were involved.

The dataset clinician labels per task are depicted in Figure 4.7. Here is noticeable that the data distribution is imbalanced in all tasks. It is observable that the trials with more observations tend to be patients with higher FAS scores (i.e., higher quality of movement). Whereas particularly, for Task 7, it was observed a generally low number of occurrences due to the task including only one trial per assessment. The above-imbalanced problem had to be solved; this is relevant because training an algorithm on this dataset would lead to poor generalization and overfitting of some scores. The ADASYN resampling method [64] was of particular interest to solve this imbalance and was applied to the dataset. Figure 4.8 depicts this technique used in Task 1, where scores 1,3, and 5 were resampled. The lower barplot shows the resampled dataset with real and synthesized points; the general observation here is that the distribution is approximately flat, meaning point distribution per class seems to be according to expected while applying ADASYN. Nevertheless, it is important to observe how the new synthesized features are distributed in a data cloud - Figure 4.9 exhibits this point clearly. It is interesting to observe that the colors of each FAS class have a clustering pattern from left to right of the X axis, where lower scores are grouped on the left and higher scores on the right—regardless, its interesting to see that the clusters are not cleanly delimited in the cloud space. In general, there is a hierarchical trend of scores. Moreover, the resamples scores seem to blend well with the real points. Figure 4.10 was developed to investigate this point further per resampled score. By analyzing the Figure, it is clear that in a generalized way, the synthesized features seem to capture the movement features according to the real points.

Next, for dimensionality reduction purposes was necessary to investigate the optimal method to perform feature selection. The results are expressed in Table 4.1 where it is noticeable that when applying a RF Regression model for all tasks, the RF built-in feature importance shows to reduce the RMSE, improve accuracy and F1 score while working to reduce feature dimensionality. When the RF method is compared to MRMR, the mean RMSE increases by 19%, the accuracy and F1 decrease respectively by 14% and 18%; for RF vs. ReliefF, the RMSE grew by 26%, the accuracy and F1 score were 14% and 20% lower; the RF vs. NCA, lead to a growth of 16% in the RMSE, a decrease of 12% and 19% in accuracy and F1 score. Thus, in general, it's favorable to consider the RF built-in feature importance method, which was indicative of providing closer estimates regarding the simulated and observed data.

When analyzing the top features selected as predictors of Model 3, the example depicted in Figure 4.11 stresses that the wrist (i.e., UE distal segment) and upper-arm (i.e., UE proximal segment) sensors are of high relevance. It was noted that the top features for all tasks primarily related to the signal peak number, which is related to the smoothness of the movements, more peaks observed indicate higher movement intensity variability. Similarly, the importance given to the energy around the dominant frequency marks the periodicity of the movement; if more of the energy is around the dominant frequency, the more periodic will be the expected movement. Moreover, correlations between the chest and the wrist sensors are necessary to interpret the reaching movement needed to grasp objects and successful task performance, as well as identifying seating position tasks as the chest sensor acts as a reference. The sensors' magnitude also appears to be likewise of larger predictive pertinence for single task predictions primarily due to considering the three-axis acceleration contributions.

Our findings described in Tables 4.2 to 4.9 show the relevance of the Hierarchical Model developed to improve the FAS single tasks estimation.

The results for Model 1, with all FAS scores, depicted in Figures 4.12 to 4.26 illustrate the Regression vs. Classification representation for this model. In general, for Regression (see Figures 4.12(a) to 4.26(a)), there is usually a trend from lower to higher scores, with over-estimation, focused on the lower scores. The confusion matrices regarding the Classification approach also portray this tendency (see Figures 4.12(b) to 4.26(b)). When comparing the tables for Model 1 with all tasks scores from 1 to 5, and comparing Model 2 with merged lower scores 1 and 2, not much difference was noted in the overall performance of the model. For that case, the model overestimated the higher scores and underestimated the lower scores. Nevertheless, when comparing Model 1 to Hierarchical Model 3, a drop in the RMSE was observed when a binary Classification was utilized in step 1, and Regression was used in step 2. Our interpretation is that first, with a Regression approach, there is a transition from a discrete value scale which classification imposes, to a continuous scale, making the decision boundaries more flexible, which is advantageous for smaller datasets like ours. Regarding the Hierarchical Model, the results support our *rationale* that an algorithm divided into two parts, including a binary Classifier for lower scores first and a certain step 2 focusing on higher scores patients enables the achievement of better performance. This high-level estimation preference is related to the clinical status of the patients who have a broad capability to benefit more from the motor evaluation by generating physiological patterns of motion of particular interest for a rehabilitation treatment adapted over the intervention care.

Analyzing all models on a task by task basis is desirable to understand the motor performance involved in each task [29] and how it is related to the results.

Initiating the analysis with Task 1, opening and filling a bottle, a bimanual task involving grasping and manipulating objects. One hand will cork the bottle, and the other will hold the bottle. The general observation here is that when there is a considerable number of trials, as in this case (3 trials), the models have a higher number of observations to train the model. Regarding Model 1, with five scores, there is a high RMSE for both RF approaches; Figure 4.12 in particular for Regression, depicts that score 1 is being overestimated, as noted in Figure 4.7 score 1 only has eighteen examples of training the model. Furthermore, Figure 4.8 reveals that forty-two synthesized points were generated for this particular score. Ideally, the real data points should be over the number of necessary synthesized points. It is important to note that in patients assessed with lower scores, there is more movement variability, in the sense that the way to perform the task in a proper manner to achieve the final goal, when a subject is more impaired, will vary more across subjects. Thus, in the lower scores, it starts deviating from the ideal movement execution, and there is much more freedom. Consequently, this requires many more points to understand and learn how to properly define the data cluster. Nonetheless, the expected movement

performance for higher scores is similar for subjects. So even with few data points, it is possible to get an approximate estimation that meets the Clinician's True one, as is noted in this case. The Hierarchical Model introduces the clinical relevance for this evaluation type in relatively high-functioning individuals. Whereas the individuals with lower FAS scores (i.e., lower functioning) are mostly not going to use their stroke/TBI affected limb during the performance of ADL, translating into having a gross measure of the FAS for that population. Thus, for further analysis, scores 1 and 2 were initially merged into Model 2 (Figure 4.12). As seen in Table 4.2, Model 2 Regression RMSE, accuracy, and F1 score improve when compared to Model 1 (-15% RMSE, +9% accuracy, +11% F1). However, an exploratory Hierarchical Model 3 was developed to take into consideration our focus on the prediction of higher FAS scores. In this model (Figure 4.13), it is clear that the binary Classification provides improved accuracy of 0.39 and an RMSE of 0.86. The Regression model was optimal in step 2; it is according as expected; regression models are preferred due to the nature of our output. The Regression plot depicts a crescent trend in the mean estimations per score with an RMSE of 0.39 and high accuracy of 0.86, as well as a 0.85 F1 score (Model 2 vs. 3 performance has -49% RMSE, +8% accuracy, +2% F1, see Table 4.2).

Considering Task 2, brushing the hair, includes three trials performed with the affected upper-limb. It is a unimanual task that involves gross motor control and grasping. For Model 1, it is notable that there is a well-defined trend in the Regression approach boxplots, with again an overestimation of score 1 (Figure 4.14(a)). There is a high variance for score 3 points estimation, with two well-defined outliers. This task movements involve deviating from the natural resting position of the forearm, so it requires patients to sustain and control a grip force during grasping and lifting the object, which is a common problem for these patients because the force produced during gripping can be slow to build up and difficult to stabilize at the required level for a specific task. So, a high movement variability is expected even for higher scores. When comparing Model 1 and 2 Regression it is observed a slight general improvement in the performance for accuracy and F1 score, while the RMSE increases (+21% RMSE, +2% accuracy, +5% F1) (see Table 4.3). Model 3 presents a slight overestimation of score 3, with an underestimation of scores 4 and 5; this can be explained due to the above-mentioned (Figure 4.15). Generally, Model 3 enables the achievement of a 0.32 RMSE, 0.68 accuracy, and 0.69 F1 score (Model 2 vs. 3 performance has -58% RMSE, +10% accuracy, +11% F1, see Table 4.3).

Further, Task 3, putting on and taking off the pen cap, is a bimanual task that concerns fine motor control manipulation. For Model 1, it is noticed that there are no true positives instances for score 1 (Figure 4.16(b)), and this is expected because there are only five real data points for this score (Figure 4.8), so the point is that even if there are methods employing dataset rebalancing, that rebalancing cannot fix the fact that there are few observations. So the algorithm rebalances the score but is still rebalancing on a limited number of subjects, especially for score 1, so it will create a compact cluster that doesn't specifically represent the lower scores. When resembling Model 1 with Model 2 Regression there is a slight decrease in performance (+7% RMSE, -2% accuracy, +20% F1, see Table 4.4). Model 3 initial binary Classification presents a good amount of true positive estimates, whereas the Regression shows that the extreme scores are tendentially over and underestimated with low variability (Figure 4.17). The overall Model 3 presents an RMSE of 0.39, accuracy of 0.61, and F1 score of 0.60 (Model 2 vs. 3 performance has -47% RMSE, +3% accuracy, +2% F1, see Table 4.4).

Task 4 is unfolding a towel, a bimanual task that affects fine specialized motor control of the hand. As depicted in Figure 4.7, this particular task can be interpreted as presenting more performance difficulty for patients, where the highest occurrences are for score 2. For Model 1, it stands out that in the Regression approach, there is not a well-defined trend in the estimation (Figure 4.18(a)). Again, there are fewer data points in score 1 (Figure 4.8). This task was mainly challenging for patients due to involv-

ing reaching and grasping with high manipulation of the towel. The general observation here is that the algorithm performs well when there are a good number of trials. Conversely, when the algorithm doesn't have enough examples of training properly, it starts dropping below a certain level. Overall, comparing Model 1 and 2 Regression presents similar performance as Task 3 (-7% RMSE, +12% accuracy, +31% F1, see Table 4.5). Finally, Model 3, enables the optimal performance with an RMSE of 0.30, accuracy of 0.68, and F1 score of 0.61 (Figure 4.19). Model 2 vs. 3 performance has -63% RMSE, +19% accuracy, +11% F1 score (see Table 4.5). Moreover, this task achieved the lowest Hierarchical Model RMSE.

Overviewing the findings for Task 5, ironing, is a unimanual task that implicates forward and backward hand movement. As observed in Model 1, the Regression provides a well-defined trend (Figure 4.20(a)); the performance of this task was of lower difficulty, providing smoother motor patterns. One observation is that these clinical scores are discrete. Still, these boundaries are removed for the Regression algorithm, so it is normal that there is some spread across the different clinical predicted scores. In the boxplots, if the predicted scores were rounded before plotting, the result for this task was potentially similar to a perfect trend line. The comparison of Model 1 and Model 2 shows a general improvement for Regression (-27% RMSE, +22% accuracy, +17% F1, see Table 4.6). As observed by the Model 3 results, there is a notable high variability for score 4 assessment (Figure 4.21(b)). Regardless, the best performance was again with a Hierarchical Model with RMSE of 0.38, an accuracy of 0.62, and F1 score of 0.60 (Model 2 vs. 3 performance has -30% RMSE, -14% accuracy, -12% F1, see Table 4.6).

Correspondingly, Task 6, lifting a box, is a bimanual task that involves reaching and grasping movement. For this task, it is interesting to acknowledge that some patients have deficits in inter-joint coordination. These deficits are evident during the reaching action, so some patients showed difficulty making smooth, continuous, and accurate arm movements in these tasks. For Model 1, there is notable poor performance; even if this task contains the highest trial number (6 trials), still the inter-joint coordination necessary for this movement seems to affect the results, which depict high variability for higher scores and a general overestimation for scores 1 and 2 (Figure 4.22). Here, Model 1 vs. Model 2 Regression show lower performance for Model 2 (+3 % RMSE, -4% accuracy, +21% F1, see Table 4.7). Model 3 helps with the latter, where lower scores are well classified in the binary Classification and higher scores focusing on the boxplot show a considerable number of observations around the mean value (Figure 4.23). Consequently, the highest RMSE was noted for this task compared to all tasks with Model 3 achieving an RMSE of 0.39, accuracy of 0.61, and F1 score of 0.58 (Model 2 vs. 3 performance has -43% RMSE, +11% accuracy, +14% F1, see Table 4.7).

Furthermore, Task 7, erasing the board, is a unimanual task that involves arm lifting and grasping movement. Here it was observed a generalized high variability in the FAS estimation. This task was performed only once with the hemiparetic side of the upper-limb, representing the task with fewer total observations. Particularly, there are few data points in all scores from 1 to 5, with score 2 having eight observations and score 5 only seven. Thus, being considered likewise a highly imbalanced dataset (Figure 4.8). Accordingly, Models 1 and 2 reflect this, with the Model 1 Regression approach not representing a defined trend in the predicted scores (Figure 4.24(a)). Interestingly the contrast between Model 1 and 2 was more noticeable for this task with a great performance improvement for Model 2 Regression (-11 % RMSE, +72% accuracy, +78% F1, see Table 4.8). A potential *rationale* is the relevance of a Hierarchical Model in small datasets where a bigger effort is made to derive higher scores with the second step of the Hierarchical Model. Figure 4.25, depicts Model 3 few observations present in this task, the RMSE was 0.36, accuracy of 0.61, and F1 score 0.58 (Model 2 vs. 3 performance has -57% RMSE, +22% accuracy, +21% F1, see Table 4.8).

Interpreting now Task 8, opening the door, is a unimanual task that involves arm lifting and grasping

movement. Model 1 depicts a general high variability with two well-defined outliers for score 5 (Figure 4.26(a)). Again, score 1 has few data points; hence the tendency is to have an overestimation of the model for these observations. When compared to Model 2, Model 1 Regression showed a lower performance (-26 % RMSE, +6% accuracy, -8% F1, see Table 4.9). Figure 4.27 represents Model 3, where score 3 is overestimated even with lower variability. Model 3, displayed higher performance with a RMSE of 0.34, accuracy of 0.64, and F1 score of 0.64 (Model 2 vs. 3 performance has -51% RMSE, -4% accuracy, +8% F1, see Table 4.9).

The FAS aggregation was performed to analyze the overall performance of Model 3. A fitted linear Regression model was used to analyze the optimal tasks for different combinations (Table 4.10). The coefficient of determination, R^2, was used as the standard metric to evaluate the final model due to its relevance in providing information regarding the model fitness of the predicted observations compared to the clinician labeled. Contextually, the optimal R^2 achieved was 0.91, indicating an excellent fit, meaning that the predicted scores are close to the Clinician's True ones (Figure 4.28). The interpretation of Table 4.10 indicates that all tasks are relevant for the optimal fit of the data; this is expected due to each task being specifically chosen by the clinicians due to its upper-limb motor importance to assess FAS particularly (on a trial by trial basis). Moreover, Tasks 1, 3, 5, and 6 revealed a model fit R^2 of 0.84, indicating that they are especially relevant contributors. In addition, Task 1 was of particular difficulty performance to the patients (see Figure 4.7), whereas Task 5 was easier to perform, as revealed by higher scores observed in the dataset; finally, Tasks 3 and 6 are indicative of mid difficulty. Thus, overall this relevant subset comprises all levels of motor quality applicability, including bimanual and unimanual manipulation of objects.

Figure 4.28 supports the previous analysis, illustrating the linear fit with all eight tasks. Table 4.11 reveals each individual task weight estimate for the final fit. Likewise, it is interesting to identify that Tasks 1, 3, 5, and 6 exhibit the higher coefficient term corroborating the higher importance for the final prediction and showing an identified statistically significant correlation with the Clinician's True scores. Still, the linear fit is projected to be higher in lower scores (i.e., overestimation) and lower in higher scores (underestimation), by remembering Figure 4.2 where it's marked that the Post-Treatment was successful for both ABI groups, being close to 4 points on average, it was likely to observe a shift towards higher scores estimates. The bias analysis was indicative that the points fell in an error range of [-11,+12]% points, meaning [0.55,0.60] FAS points (Figure 4.29). Thus, an error of fewer than 1 FAS point can be achieved in the worst-case scenario. Altogether, this suggests that the model's bias was negligible and represents suitable automation of the FAS scores.

5.1.2 FMA Estimation

Next is accomplished the interpretation of the derivation of the widely used FMA scale to evaluate the severity of upper-limb motor impairments. The clinicians administered the FMA without the use of wearable sensing technology; its total score is reported in items of clinical relevance. As opposed to the FAS clinical scale, which was assessed per task trial, the FMA scale is assessed once per subject visit (i.e., Baseline and Post-Treatment). Previous studies reported the feasibility of deriving the FMA from wearable sensors data collected from the WMFT [53]. Our analysis concentrates on incorporating not only the wearable sensor data but also taking advantage of the Hierarchical Model developed to derive FAS predictions in order to improve the final FMA model predictions.

The analysis of the direct correlation between the Clinician's Predicted FAS and FMA scores revealed a positive association between the scales, indicative that there is indeed potential to use the FAS

predictions as predictive information of the FMA (Figure 4.30). The dataset clinician labeled is shown in Figure 4.31, where is likewise noted an imbalanced distribution of points; the proximal scores were generally lower than the distal segment scores. However, the total FMA is marked by the variability of the distribution of the scores with a minimum score of 21 points (i.e., below 21 points, there are no occurrences). The ADASYN method was desirable to apply in this case. Hence, the scores were grouped for resampling purposes, shown in Table 4.12 whose distribution can be found highlighted in Figure 4.32, reveals that in general, the total score shows a better-uniformed points distribution. The results from the ADASYN method (Figure 4.33) were also reasonable for all tasks.

A model was developed without the predictive information of the FAS as an initial approach. Soon, it was clear that the FAS introduced model improvements, as shown in Figure 4.34 from Model 2. It was noted that the RF gave nonzero importance to the FAS prediction feature; moreover, this feature ranked top three for all tasks (Table 4.13). The latter is a good indication of the feasibility of this analysis. Additionally, it is interesting to interpret how the upper-arm features are amongst the most important. The upper-arm sensor, being positioned in a body segment with a lower range of motion than the wrist, generally includes smoother movements and, therefore, less variability. Thus, the affected side's upper-arm speed magnitude was highly important. The same occurs for the chest sensor, which is mainly used as a reference importance being as well as a helpful indicator if the patient is in a sitting position, thus highlighting reaching action movements (particularly, movement periodicity); moreover, other features, such as the correlation between the chest and wrist units, were identified as relevant due to the chest sensor constituting a good reference point between the time series. The affected wrist features were also of relevance for all tasks. Furthermore, the number of peaks in the signal is related to the smoothness of the movements; fewer peaks in the signal revealed smoother motions.

The overall error estimation varies for each patient (Figure 4.36), around 6 points per single task, indicating that Model 2 worked for this estimation globally (Figure 4.35).

An analysis was performed to check the subset of tasks that were more representative to estimate the FMA. The model to aggregate the FMA scores revealed that Tasks 5 (Ironing), Task 4 (Unfolding Towel), and Task 1 (Opening/Filling a Bottle) were optimal predictors of the model (Figure 4.37). The results are intriguing; Task 5 was noted as one of the most straightforward tasks to perform in the FAS, while Tasks 1 and 4 were considered the most challenging. Since it was observed in Task 5 that patients had more quality of movement, it is a representation that the outcomes include the FAS factor relevance. Then, the fact that the most challenging tasks were considered significant predictors makes sense since the FMA assesses the severity of motor impairments, and the motor performance patterns vary more in more challenging tasks, capturing better the seriousness of the hemiparetic impairments.

Figure 4.38 reveals a steady overestimation of lower scores and underestimation of higher scores. When compared to the FAS linear fit, the trend is similar but more well noted for the FMA, where the coefficient R^2 with the optimal subset of tasks was 0.79, indicating a good positive fit, meaning that the predicted scores are close to the Clinician's True one. Likewise, bias analysis (Figure 4.39) was indicative that the points fell in an error range of $\pm$ 12 points (i.e., confidence interval) in the worst-case scenario. The bias was minimal (i.e., close to zero), which denotes a good sign, meaning that when comparing the two measures, the error is within the confidence interval and is not expected to be statistically significant. Due to the FMA being an imbalanced dataset with 66 points scale as opposed to the FAS with 5 points, it was expected the remarked higher bias between the FMA scores. Furthermore, Figure 4.40 denotes the FMA scores estimation error analyzed as a function of the Clinician's True FMA, where it was noted that the estimation error is overestimated for lower scores and underestimated for higher scores; still, for most cases is included in the Minimal Detectable Change (MDC) boundaries.

5.2 Study Limitations

The study encompasses some limitations that should be taken into account when considering our analysis approach and final results.

Firstly, the cohort of ABI patients constituted a heterogeneous but small sample size. Therefore, with an additional enrolled number of patients, it would be expected that the overall model performance would improve by providing the ML algorithms with more training observations. Moreover, the dataset was imbalanced, meaning that the quality of movement and severity of impairments of the patients was unevenly distributed. Thus, the synthetic point generation methods worked to solve this imbalance. However, as observed, some clinical scores had few observations; therefore, even if there were newly synthesized observations, they were generated on a small observation set which doesn't necessarily represent the score cluster. Furthermore, the synthesized points are redundant in the sense that they do not have the same informative power as the real clinician-assessed data.

It is also pertinent to acknowledge that at the time of the data acquisition, solely the accelerometers tool was used from the IMU; this choice was based primarily on the fact that the aforesaid enables less power consumption with a reduction of the data processing complexity. Consequently, the span of the experiment could be agreeably extended; hence data could be acquired for a longer period of time. Nevertheless, new devices have improved IMU power durability. Accordingly, would be interesting to have gyroscope information regarding the orientation of the rotation of the upper-arm segments.

The sensor calibration was employed to understand the patients' motor performance and relate the sensor's orientation to the individual sensors' axis movement in this study. To this end, it was required to estimate each sensor offset due to the sensor's position being altered in each assessment. The developed offset estimation method introduces a small noise source as it was considered that the resting period condition on which the offset was estimated was equivalent to 1g. However, the aforementioned would be less relevant to compensate for if the same sensor set was placed invariably in the exact positions.

Regarding the labels used as the "true" scores for the FAS and FMA, it is interesting to note that the clinicians responsible for labeling the patients were not consistently the same per assessment. Even though the clinical staff has strong clinical training to use the clinical scales, there should still be some inter-rater reliability (i.e., the overall reproducibility of scores between two reviewers), generally negligible. A possible approach would be to ask different clinicians to assess the scores for the Baseline and Post-Treatment assessments to ensure the clinical labels are consistently generated.

Additionally, even though a RF approach brings remarkable advantages for our small cohort, they arrive at the expense of complexity, meaning there is a loss of model interpretability, particularly for the developed models. Furthermore, the number of trees to include in the RF model was selected, taking into account that the increase in the number of trees introduces more computation cost needs (i.e., higher memory and training time), so a trade-off had to be evaluated. Furthermore, I believe that it would be interesting to test the algorithms developed in this Dissertation on a new cohort that followed the same study inclusion criteria defined in the Subsection 3.1.2.

Further, due to the small dataset, which is prone to model overfitting, the LOSOCV approach, including a leaving-one-subject out technique to validate the model, was employed given the proven efficiency of this approach for a low number of observations [75].

Finally, the assessments took part at two time points; the ideal would be to assess patients consistently over the upper-limb rehabilitation process to frequently analyze the outcomes and accordingly adapt the plan of care. Moreover, the ADL tasks were performed in a scripted clinical setting; it would be highly desirable to transition to a home setting for patients to feel familiarized with the community scenarios.

Chapter 6

Conclusions and Future Work

The main results of this study are highlighted in this chapter, which likewise provides the contributions of this study research and recommendations for future research in upper-limb-focused motor rehabilitation of Acquired Brain Injury patients.

6.1 Conclusions

Clinical score estimates were derived from wearable sensors technology data in Acquired Brain Injury patients to track motor changes in response to a clinical intervention focused on upper-limb rehabilitation. The objective of this project was twofold: primarily, the aim was to develop dedicated ML algorithms, including new methodologies to estimate the Functional Ability Scale concerning the patient's quality of movement during the performance of ADLs; secondly, the focus was to develop extended ML algorithms to derive the Fugl-Meyer Assessment, for the severity of motor impairment assessment, taking advantage of the knowledge learned from the preceding Functional Ability Scale estimations. With this, clinicians are empowered with tools to consistently perform rehabilitation interventions in the plan of care, adjusted accordingly to the outcomes of patient-specific responsiveness and in an unobtrusive way.

As mentioned, following an event such as a TBI or stroke, clinical assessments are essential to evaluate the rehabilitation intervention aiming to treat the sequelae of these events, specially upper-limb hemiparesis, which constitutes a significant limitation in the activities of daily living performance. However, traditionally the rehabilitation assessment outcome measures are solely collected at two time points: baseline and discharge. The clinicians usually perform the assessments in order to identify movement dysfunctions. The latter is of highly challenging reproducibility as each clinician has a particular background, clinical judgment, and time available. Consequently, it is desirable to objectively quantify the motor improvement or decline in motor gains.

This Dissertation leveraged the potential shown by wearable sensors and powerful Machine Learning algorithms, which constitute technologies with the capacity to enable researchers and clinicians to collect data at multiple time points and consequently make the clinical intervention adjustment less challenging and more efficient. Accordingly, with our proposal, automation of two widely used clinical scores is achieved with the drive to empower patients to regain confidence in ADL performance by maximizing the motor gains over the care plan.

Thus, to assess a patient's responsiveness to clinical intervention, the FAS and FMA scales were selected due to their well-known validity, reliability, and responsiveness in assessing upper-limb movements from IMU sensors [48],[49]. The models employed are based on a RF Ensemble Learning method-

ology which is seen as one of the most powerful ML available [38] and works by aggregating a group of decision tree predictors (i.e., an ensemble).

Broadly translated, our findings indicate that the analysis for the movement quality, FAS, revealed that when dealing with the small imbalanced dataset, an ADASYN rebalancing method resulted efficient. Furthermore, a dedicated Hierarchical model with a Binary Classifier and Regression on the second level is the best solution to derive scores with a particular focus on high-motor functioning patients. The final model achieved excellent performance with a coefficient of determination, R^2, of 0.91.

In addition, the main contributions of this research pertaining the motor Impairment, FMA, highlight that the subset of Tasks 1 (Opening/Filling a Bottle), Task 4 (Unfolding Towel), and Task 5 (Ironing) emerges as the optimal subset of tasks to derive the FMA. Moreover, adding the FAS scores into the final model for both single tasks and final combined predictions was revealed to be key to improving the performance by providing quality movement knowledge. As a result, the final model achieved good performance with a coefficient of determination, R^2, of 0.79.

Previously published studies have shown the feasibility of deriving the FMA from wearable sensors data collected from the WMFT [53]. Therefore, our analysis is concentrated on incorporating not only the wearable sensor data but also taking advantage of the hierarchical model developed to derive FAS predictions in order to improve the final FMA model predictions. Overall, our results are encouraging, demonstrating the feasibility of this analysis with relevant outcomes.

The performing tasks were also explicitly selected and adapted to reproduce motor patterns of activities of daily living. Thus, in this research, the patients were assessed on tasks that enabled more freedom of upper-limb movements, concentrating on successfully accomplishing each task goal.

Finally, as demonstrated, the analysis conducted in this research exhibited the relevance of developing improved methodological quality for assessing ABI patients' upper-limb motor performance in this area, showing the feasibility of these procedures to assist clinicians in overcoming patient's motor nonuse, improve the hemiparetic upper extremity range of motion and ADL task performance by performing precision-rehabilitation.

6.2 Contributions

As a final remark, I want to highlight that the results of this research project have been used to contribute to the IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE), which took place in Washington, D.C., USA, on November 17-19, 2022 (Appendix A).

- **C. Ernesto**, F. Parisi, C. Adans-Dester, A. O'Brien, G. Vergara-Diaz, R. Black-Schaffer, R. Zafonte, H. Ferreira, and P. Bonato, "Wearable Technology and Machine Learning to Monitor Upper-Limb Use in Brain Injury Survivors," in *2022 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*. Arlington, VA, USA: Institute of Electrical and Electronics Engineers Inc., 2022, pp. 180-181.

Furthermore, the results presented in this Master's Dissertation are foreseen to be published in one peer-reviewed publication.

6.3 Future Work

This work portrays the potential of using wearable sensors and powerful Machine Learning algorithms to improve the assessment of upper-limb rehabilitation outcomes post Acquired Brain Injury lesions. Innovative tools were developed to tackle patient-specific motor gain rehabilitation driven by quantitative measures, empowering clinicians and patients and translating into shorter, more consistent rehabilitation programs. Similarly, these technologies show the potential to diminish the financial burdens associated with stroke and TBI lesions, which represent a significant drain on economic resources, affecting society and health services.

The Motion Analysis laboratory is currently conducting a study that aims to collect data using IMU sensing technology in stroke patients with upper-limb hemiparesis. The study assessment brings more attention to the quality recording of the signals and environmental conditions. Furthermore, more subjects were recruited and enrolled in the study to heighten the statistical power of the analysis. The study is encouraging due to transitioning from a clinical setting, such as the one included in this Dissertation, into a simulated apartment located at the Spaulding Rehabilitation Hospital. The activities are not scripted, enabling an even less constrained environment to assess upper-limb motion. Again, the sensing technology is innovative, including utilizing two ring sensors per hand to replace the two-axis accelerometers used solely in the patient's affected index and thumb. The hypothesis is that this new technology can better capture fine motor control when manipulating objects involved in the activities. I was able to conduct the first data collection and several upcoming ones, and it was remarkable to observe the patient's activities of daily living performance. Also, in this case, there was a setup of video recordings, enabling multiple clinicians to label the assessment a posteriori with the possibility of rewatching each clip, hence enabling less interrater variability. After listening to the user requirements, I developed the templates of the clinical labeling system to be used in the novel study. It is interesting to note that with this new clinical study setup, more clinical scales can be labeled due to the overcoming of the limited time available in the patient visit; in this way, new algorithms can be created, and hopefully, more clinical scales can be automated and improved.

Furthermore, it would be desirable to assess not only patients' motor capacity (evaluated in the clinic) but likewise the patient's performance which is what patients can actually do in the home and community settings. This is particularly relevant because ABI patients usually use compensatory strategies in the home setting (i.e., have the motor ability but don't particularly use it with the hemiparetic limb), making it of high relevance to provide constant feedback during the plan of care. Thus, the Motion Analysis team is collaborating and hosting workshops in Connected Eldercare with other US universities as part of the NSF Engineering Research Center. This project aims to discuss and evaluate pilot studies, including the use of wearable sensing technology in the home and community settings so that clinicians and caregivers can access elderly homes and provide care services remotely.

Regarding the sensing technology, overall, the wearable sensors used in this Dissertation met the user requirements such as accuracy, comfort, and ease of use. Advances in micro-electromechanical systems (MEMS) have facilitated the development of miniaturized accelerometers; this is of special relevance because, in-home and community settings, the long-term monitoring of patients implies that patients or caregivers would have to place the sensing technology in the upper-limbs. Thus, exploring the possibility of a minimum set of sensors to be employed should be leveraged and studied.

To conclude, in the future, novel wearable devices to monitor ABI patients outside of the clinical setting will enable a transition from clinician-reported outcomes to sensor-based outcome estimates that, among others, maximize Acquired Brain Injury patients' motor gains.

Bibliography

[1] P. Raghavan, "Upper Limb Motor Impairment After Stroke," *Physical Medicine and Rehabilitation Clinics of North America*, vol. 26, no. 4, pp. 599–610, 2015. [Online]. Available: http://doi.org/10.1016/j.pmr.2015.06.008

[2] "WHO — International Classification of Functioning, Disability and Health (ICF)," *WHO*, 2019, (Last Accessed on 12/04/2022). [Online]. Available: http://www.who.int/classifications/icf/en/

[3] C. Adans-Dester, S. E. Fasoli, E. Fabara, N. Menard, A. B. Fox, G. Severini, and P. Bonato, "Can kinematic parameters of 3D reach-to-target movements be used as a proxy for clinical outcome measures in chronic stroke rehabilitation? An exploratory study," *Journal of NeuroEngineering and Rehabilitation*, vol. 17, no. 1, pp. 1–13, 2020. [Online]. Available: http://doi.org/10.1186/s12984-020-00730-1

[4] P. MacEira-Elvira, T. Popa, A. C. Schmid, and F. C. Hummel, "Wearable technology in stroke rehabilitation: Towards improved diagnosis and treatment of upper-limb motor impairment," *Journal of NeuroEngineering and Rehabilitation*, vol. 16, no. 1, pp. 1–18, 2019. [Online]. Available: http://doi.org/10.1186/s12984-019-0612-y

[5] Q. Wang, P. Markopoulos, B. Yu, W. Chen, and A. Timmermans, "Interactive wearable systems for upper body rehabilitation: A systematic review," *Journal of NeuroEngineering and Rehabilitation*, vol. 14, no. 1, p. 20, 2017. [Online]. Available: http://doi.org/10.1186/s12984-017-0229-y

[6] C. Pao, "Dynamic Calibration for IMU Motion Sensors - CEVA's Experts blog," 2019, (Last Accessed on 20/05/2022). [Online]. Available: https://www.ceva-dsp.com/ourblog/optimize-your-imu-with-dynamic-calibration/

[7] C. Adans-Dester, N. Hankov, A. O'Brien, G. Vergara-Diaz, R. Black-Schaffer, R. Zafonte, J. Dy, S. I. Lee, and P. Bonato, "Enabling precision rehabilitation interventions using wearable sensors and machine learning to track motor recovery," *npj Digital Medicine*, vol. 3, no. 1, p. 121, 2020. [Online]. Available: http://doi.org/10.1038/s41746-020-00328-w

[8] S. Del Din, S. Patel, C. Cobelli, and P. Bonato, "Estimating fugl-meyer clinical scores in stroke survivors using wearable sensors," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, Boston, MA, USA, 2011, pp. 5839–5842. [Online]. Available: http://doi.org/10.1109/IEMBS.2011.6091444

[9] L. Yu, D. Xiong, L. Guo, and J. Wang, "A remote quantitative Fugl-Meyer assessment framework for stroke patients based on wearable sensor networks," *Computer Methods and Programs in Biomedicine*, vol. 128, pp. 100–110, 2016. [Online]. Available: http://doi.org/10.1016/j.cmpb.2016.02.012

[10] J. See, L. Dodakian, C. Chou, V. Chan, A. McKenzie, D. J. Reinkensmeyer, and S. C. Cramer, "A standardized approach to the fugl-meyer assessment and its implications for clinical trials," *Neurorehabilitation and neural repair*, vol. 27, no. 8, pp. 732–741, 2013. [Online]. Available: https://doi.org/10.1177/1545968313491000

[11] 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," *Archives of physical medicine and rehabilitation*, vol. 82, no. 6, pp. 750–755, 2001. [Online]. Available: https://doi.org/10.1053/APMR.2001.23183

[12] 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," *Neurorehabilitation and neural repair*, vol. 16, no. 3, pp. 232–240, 2002. [Online]. Available: https://doi.org/10.1177/154596802401105171

[13] V. Y. Ma, L. Chan, and K. J. Carruthers, "Incidence, Prevalence, Costs, and Impact on Disability of Common Conditions Requiring Rehabilitation in the United States: Stroke, Spinal Cord Injury, Traumatic Brain Injury, Multiple Sclerosis, Osteoarthritis, Rheumatoid Arthritis, Limb Loss, and Back Pain," *Archives of Physical Medicine and Rehabilitation*, vol. 95, no. 5, pp. 986–995, 2014. [Online]. Available: http://doi.org/10.1016/j.apmr.2013.10.032

[14] E. Benjamin, P. Muntner, A. Alonso, M. Bittencourt, C. Callaway, A. Carson, A. Chamberlain, A. Chang, S. Cheng, S. Das, F. Delling, L. Djousse, M. Elkind, J. Ferguson, M. Fornage, L. C. Jordan, S. Khan, B. Kissela, K. Knutson, T. Kwan, D. Lackland, T. Lewis, J. Lichtman, C. Longenecker, M. S. Loop, P. Lutsey, S. Martin, K. Matsushita, A. Moran, M. Mussolino, and M. O'Flaherty, "Heart Disease and Stroke Statistics-2019 Update: A Report From the American Heart Association," *Circulation*, vol. 139, no. 10, pp. e56–e528, 2019. [Online]. Available: http://doi.org/10.1161/CIR.0000000000000659

[15] J. Lo, L. Chan, and S. Flynn, "A Systematic Review of the Incidence, Prevalence, Costs, and Activity and Work Limitations of Amputation, Osteoarthritis, Rheumatoid Arthritis, Back Pain, Multiple Sclerosis, Spinal Cord Injury, Stroke, and Traumatic Brain Injury in the United States," *Archives of Physical Medicine and Rehabilitation*, vol. 102, no. 1, pp. 115–131, 2020. [Online]. Available: http://doi.org/10.1016/j.apmr.2020.04.001

[16] R. Lemmens, A. Timmermans, Y. Janssen-Potten, R. Smeets, and H. Seelen, "Valid and reliable instruments for arm-hand assessment at ICF activity level in persons with hemiplegia: A systematic review," *BMC Neurology*, vol. 12, no. 1, pp. 1–17, 2012. [Online]. Available: http://doi.org/10.1186/1471-2377-12-21

[17] K. Salter, N. Campbell, M. Richardson, S. Mehta, J. Jutai, L. Zettler, M. Moses, A. McClure, R. Mays, N. Foley, and R. Teasell, "EBRSR: Evidence-Based Review of Stroke Rehabilitation, 20 Outcome Measures in Stroke Rehabilitation," Tech. Rep., 2020, (Last Accessed on 10/04/2022). [Online]. Available: www.ebrsr.com

[18] M. Campbell, *Physical Management in Neurological Rehabilitation*, chapter 7: Acquired brain injury: Trauma and pathology, 2nd ed., pp. 103-124, Mosby, 2004.

[19] V. Feigin, S. Barker, R. Krishnamurthi, A. Theadom, and N. Starkey, "Epidemiology of ischaemic stroke and traumatic brain injury," *Best practice & research. Clinical anaesthesiology*, vol. 24, no. 4, pp. 485–494, 2010. [Online]. Available: http://doi.org/10.1016/j.bpa.2010.10.006

[20] C. Johnson, M. Nguyen, G. Roth, E. Nichols, T. Alam, and D. Abate, "Global, regional, and national burden of stroke, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016," *The Lancet Neurology*, vol. 18, no. 5, pp. 439–458, 2019. [Online]. Available: http://doi.org/10.1016/S1474-4422(19)30034-1

[21] M. Dewan, A. Rattani, S. Gupta, R. Baticulon, Y. C. Hung, M. Punchak, A. Agrawal, A. Adeleye, M. Shrime, A. Rubiano, J. Rosenfeld, and K. Park, "Estimating the global incidence of traumatic brain injury," *Journal of Neurosurgery*, vol. 130, no. 4, pp. 1080–1097, 2019. [Online]. Available: http://doi.org/10.3171/2017.10.JNS17352

[22] J. Elkins and C. Johnston, "Thirty-year projections for deaths from lschemic ischemic stroke in the United States," *Stroke*, vol. 34, no. 9, pp. 2109–2112, 2003. [Online]. Available: http://doi.org/10.1161/01.STR.0000085829.60324.DE

[23] D. Morris, G. Uswatte, S. Mckay, and J. M. Meythaler, "Constraint-induced movement therapy for recovery of upper-limb function following traumatic brain injury," *Journal of Rehabilitation Research and Development*, vol. 42, no. 6, pp. 769—778, 2005. [Online]. Available: http://doi.org/10.1682/JRRD.2005.06.0094

[24] J. Elbaum and D. Benson, *Acquired Brain Injury: An Integrative Neuro-Rehabilitation Approach*, 1st ed., pp. 18-38, Springer, 2007.

[25] P. van Vliet and M. Sheridan, "Coordination Between Reaching and Grasping in Patients With Hemiparesis and Healthy Subjects," *Archives of Physical Medicine and Rehabilitation*, vol. 88, no. 10, pp. 1325–1331, 2007. [Online]. Available: http://doi.org/10.1016/j.apmr.2007.06.769

[26] E. Miller, L. Murray, L. Richards, R. Zorowitz, T. Bakas, P. Clark, and S. Billinger, "Comprehensive overview of nursing and interdisciplinary rehabilitation care of the stroke patient: A scientific statement from the American heart association," *Stroke*, vol. 41, no. 10, pp. 2402–2448, 2010. [Online]. Available: http://doi.org/10.1161/STR.0b013e3181e7512b

[27] M. Kahn, R. Clark, K. Bower, B. Mentiplay, and G. Williams, "Measurements scales for associated reactions of the upper limb in stroke and traumatic brain injury (TBI): a systematic review," *Physiotherapy*, vol. 101, no. 1, pp. 703–704, 2015. [Online]. Available: http://doi.org/10.1016/j.physio.2015.03.3554

[28] J. Tretriluxana, J. Gordon, B. E. Fisher, and C. Winstein, "Hemisphere specific impairments in reach-to-grasp control after stroke: Effects of object size," *Neurorehabilitation and Neural Repair*, vol. 23, no. 7, pp. 679–691, 2009. [Online]. Available: http://doi.org/10.1177/1545968309332733

[29] J. Carr and R. Shepherd, *Stroke Rehabilitation*, chapter 5 - Reaching and manipulation, 1st ed., pp. 159-206, Butterworth-Heinemann, 2003.

[30] T. Nguyen, D. Stewart, P. Rosenbaum, S. Baptiste, O. Kraus de Camargo, and J. Gorter, "Using the ICF in transition research and practice? Lessons from a scoping review," *Research in Developmental Disabilities*, vol. 72, pp. 225–239, 2018. [Online]. Available: https://doi.org/10.1016/j.ridd.2017.11.003

[31] T. Threats, "Use of the ICF for Guiding Patient-Reported Outcome Measures," *Perspectives on Neurophysiology and Neurogenic Speech and Language Disorders*, vol. 22, no. 4, pp. 128–135, 2012. [Online]. Available: http://doi.org/10.1044/nnsld22.4.128

[32] T. Watanabe, "Meaningful Assessment in Patients with Acquired Brain Injuries," *Physical Medicine and Rehabilitation Clinics of North America*, vol. 29, no. 3, pp. 437–444, 2018. [Online]. Available: https://doi.org/10.1016/j.pmr.2018.04.006

[33] E. Niederberger, M. Parnham, J. Maas, and G. Geisslinger, "4 Ds in health research—working together toward rapid precision medicine," *EMBO Molecular Medicine*, vol. 11, no. 11, p. e10917, 2019. [Online]. Available: http://doi.org/10.15252/emmm.201910917

[34] T. Hulsen, S. Jamuar, A. Moody, J. Karnes, O. Varga, S. Hedensted, R. Spreafico, D. Hafler, and E. McKinney, "From big data to precision medicine," *Frontiers in Medicine*, vol. 6, p. 34, 2019. [Online]. Available: http://doi.org/10.3389/fmed.2019.00034

[35] F. Porciuncula, A. V. Roto, D. Kumar, I. Davis, S. Roy, C. Walsh, and L. Awad, "Wearable Movement Sensors for Rehabilitation: A Focused Review of Technological and Clinical Advances," *Journal of Injury, Function and Rehabilitation*, vol. 10, no. 9, pp. S220–S232, 2018. [Online]. Available: http://doi.org/10.1016/j.pmrj.2018.06.013

[36] J. Barth, J. Klaesner, and C. E. Lang, "Relationships between accelerometry and general compensatory movements of the upper limb after stroke," *Journal of NeuroEngineering and Rehabilitation*, vol. 17, no. 1, p. 138, 2020. [Online]. Available: http://doi.org/10.1186/s12984-020-00773-4

[37] C. Hughes, M. Baye, C. Gordon-Murer, A. Louie, S. Sun, G. Belay, and X. Zhang, "Quantitative Assessment of Upper Limb Motor Function in Ethiopian Acquired Brain Injured Patients Using a Low-Cost Wearable Sensor," *Frontiers in Neurology*, vol. 10, p. 1323, 2019. [Online]. Available: https://doi.org/10.3389/fneur.2019.01323

[38] A. Géron, *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*, 2nd ed., pp. 103-124, O'Reilly Media Inc., 2019.

[39] A. Fugl Meyer, L. Jaasko, and I. Leyman, "The post stroke hemiplegic patient. I. A method for evaluation of physical performance," *Scandinavian Journal of Rehabilitation Medicine*, vol. 7, no. 1, pp. 13–31, 1975. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/1135616/

[40] S. Wolf, P. Catlin, M. Ellis, A. Archer, B. Morgan, and A. Piacentino, "Assessing Wolf Motor Function Test as outcome measure for research in patients after stroke," *Stroke*, vol. 32, no. 7, pp. 1635–1639, 2001. [Online]. Available: http://doi.org/10.1161/01.STR.32.7.1635

[41] E. Taub, N. E. Miller, T. Novack, E. W. Cook, W. Fleming, C. Nepomuceno, J. Connell, and J. Crago, "Technique to improve chronic motor deficit after stroke." *Archives of physical medicine and rehabilitation*, vol. 74, no. 4, pp. 347–54, 1993. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/8466415/

[42] B. Hingtgen, J. McGuire, M. Wang, and G. F. Harris, "An upper extremity kinematic model for evaluation of hemiparetic stroke," *Journal of Biomechanics*, vol. 39, no. 4, pp. 681–688, 2006. [Online]. Available: http://doi.org/10.1016/j.jbiomech.2005.01.008

[43] C. Massie, S. Fritz, and M. Malcolm, "Elbow Extension Predicts Motor Impairment and Performance after Stroke," *Rehabilitation Research and Practice*, vol. 2011, pp. 1–7, 2011. [Online]. Available: http://doi.org/10.1155/2011/381978

[44] L. Van Dokkum, I. Hauret, D. Mottet, J. Froger, J. Métrot, and I. Laffont, "The contribution of kinematics in the assessment of upper limb motor recovery early after stroke," *Neurorehabilitation and Neural Repair*, vol. 28, no. 1, pp. 4–12, 2014. [Online]. Available: http://doi.org/10.1177/1545968313498514

[45] H. L. Chen, K. C. Lin, R. J. Liing, C. Y. Wu, and C. L. Chen, "Kinematic measures of Arm-trunk movements during unilateral and bilateral reaching predict clinically important change in perceived arm use in daily activities after intensive stroke rehabilitation," *Journal of NeuroEngineering and Rehabilitation*, vol. 12, no. 1, p. 84, 2015. [Online]. Available: http://doi.org/10.1186/s12984-015-0075-8

[46] C. C. Yang and Y. L. Hsu, "A review of accelerometry-based wearable motion detectors for physical activity monitoring," *Sensors*, vol. 10, no. 8, pp. 7772–7788, 2010. [Online]. Available: http://doi.org/10.3390/s100807772

[47] A. Moreira, S. Queirós, J. Fonseca, P. Rodrigues, N. Rodrigues, and J. Vilaça, "Real-time hand tracking for rehabilitation and character animation," in *SeGAH 2014 - IEEE 3rd International Conference on Serious Games and Applications for Health, Books of Proceedings*. Rio de Janeiro, Brazil: Institute of Electrical and Electronics Engineers Inc., 2014, pp. 1–8. [Online]. Available: http://doi.org/10.1109/SeGAH.2014.7067086

[48] B. Kopp, A. Kunkel, H. Flor, T. Platz, U. Rose, K. H. Mauritz, K. Gresser, K. L. McCulloch, and E. Taub, "The arm motor ability test: Reliability, validity, and sensitivity to change of an instrument for assessing disabilities in activities of daily living," *Archives of Physical Medicine and Rehabilitation*, vol. 78, no. 6, pp. 615–620, 1997. [Online]. Available: http://doi.org/10.1016/s0003-9993(97)90427-5

[49] F. Malouin, L. Pichard, C. Bonneau, and A. Durand, "Evaluating motor recovery early after stroke: comparison of the fugl-meyer assessment and the motor assessment scale," *Archives of Physical Medicine and Rehabilitation*, vol. 75, no. 11, pp. 1206–1212, 1994. [Online]. Available: http://doi.org/10.1016/0003-9993(94)90006-x

[50] J. H. Lin, M. J. Hsu, C. F. Sheu, T. S. Wu, R. T. Lin, C. H. Chen, and C. L. Hsieh, "Psychometric comparisons of 4 measures for assessing upper-extremity function in people with stroke," *Physical Therapy*, vol. 89, no. 8, pp. 840–850, 2009. [Online]. Available: http://doi.org/10.2522/ptj.20080285

[51] J. Van Der Lee, H. Beckerman, D. Knol, H. De Vet, and L. Bouter, "Clinimetric properties of the motor activity log for the assessment of arm use in hemiparetic patients," *Stroke*, vol. 35, no. 6, pp. 1410–1414, 2004. [Online]. Available: http://doi.org/10.1161/01.STR.0000126900.24964.7e

[52] 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," *Archives of Physical Medicine and Rehabilitation*, vol. 87, no. 10, pp. 1340–1345, 2006. [Online]. Available: http://doi.org/10.1016/j.apmr.2006.06.006

[53] S. Patel, R. Hughes, T. Hester, J. Stein, M. Akay, J. G. Dy, and P. Bonato, "A novel approach to monitor rehabilitation outcomes in stroke survivors using wearable technology," *Proceedings of the IEEE*, vol. 98, no. 3, pp. 450–461, 2010. [Online]. Available: http://doi.org/10.1109/JPROC.2009.2038727

[54] J. Wang, L. Yu, J. Wang, L. Guo, X. Gu, and Q. Fang, "Automated fugl-meyer assessment using svr model," in *2014 IEEE International Symposium on Bioelectronics and Bioinformatics*. Chung Li, Taiwan: Institute of Electrical and Electronics Engineers Inc., 2014, pp. 1–4. [Online]. Available: http://doi.org/10.1109/ISBB.2014.6820907

[55] G. Saggio, G. Orengo, A. Pallotti, V. Errico, and M. Ricci, "Sensory Systems for Human Body Gesture Recognition and Motion Capture," in *2018 International Symposium on Networks, Computers and Communications, ISNCC 2018*. Rome, Italy: Institute of Electrical and Electronics Engineers Inc., 2018, pp. 1–6. [Online]. Available: http://doi.org/10.1109/ISNCC.2018.8531054

[56] MATLAB, *version 9.10.0 (R2021a)*. Natick, Massachusetts: The MathWorks Inc., 2021.

[57] E. Wade, A. R. Parnandi, and M. J. Mataric, "Using socially assistive robotics to augment motor task performance in individuals post-stroke," in *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*. San Francisco, CA, USA: Institute of Electrical and Electronics Engineers Inc., 2011, pp. 2403–2408. [Online]. Available: http://doi.org/10.1109/IROS.2011.6095107

[58] C. Verplaetse, "Inertial proprioceptive devices: self-motion-sensing toys and tools," *IBM Systems Journal*, vol. 35, no. 3.4, pp. 639–650, 1996. [Online]. Available: https://doi.org/10.1147/sj.353.0639

[59] T. S. Patterson, M. D. Bishop, T. E. McGuirk, A. Sethi, and L. G. Richards, "Reliability of upper extremity kinematics while performing different tasks in individuals with stroke," *Journal of Motor Behavior*, vol. 43, no. 2, pp. 121–130, 2011. [Online]. Available: https://doi.org/10.1080/00222895.2010.548422

[60] B. Garrison and E. Wade, "Relative accuracy of time and frequency domain features to quantify upper extremity coordination," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Milan, Italy: Annual International Conference of the IEEE EMBC, 2015, pp. 4958–4961. [Online]. Available: https://doi.org/10.1109/EMBC.2015.7319504

[61] T. Hester, R. Hughes, D. M. Sherrill, B. Knorr, M. Akay, J. Stein, and P. Bonato, "Using wearable sensors to measure motor abilities following stroke." Cambridge, MA, USA: Proceedings - BSN 2006: International Workshop on Wearable and Implantable Body Sensor Networks, 2006, pp. 5–8. [Online]. Available: https://doi.org/10.1109/BSN.2006.57

[62] W. Dargie, "Analysis of time and frequency domain features of accelerometer measurements," in *2009 Proceedings of 18th International Conference on Computer Communications and Networks*, San Francisco, CA, USA, 2009, pp. 1–6. [Online]. Available: https://doi.org/10.1109/ICCCN.2009.5235366

[63] L. Bao and S. S. Intille, "Activity recognition from user-annotated acceleration data," in *Pervasive Computing*, vol. 3001. Springer Berlin Heidelberg, 2004, pp. 1–17. [Online]. Available: https://doi.org/10.1007/978-3-540-24646-6_1

[64] H. He, Y. Bai, E. Garcia, and S. Li, "Adasyn: Adaptive synthetic sampling approach for imbalanced learning," vol. 74, no. 4. Hong Kong: 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), 2008, pp. 1322–1328. [Online]. Available: http://doi.org/10.1109/IJCNN.2008.4633969

[65] "Standardized z-scores," (Last Accessed on 04/08/2022). [Online]. Available: https://www.mathworks.com/help/stats/zscore.html

[66] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001. [Online]. Available: http://doi.org/10.1023/A:1010933404324

[67] D. Lagomarsino, . V. Tofani, . S. Segoni, . F. Catani, and . N. Casagli, "A tool for classification and regression using random forest methodology: Applications to landslide susceptibility mapping and soil thickness modeling," *Environmental Modeling Assessment*, vol. 22, p. 201–214, 2016. [Online]. Available: http://doi.org/10.1007/s10666-016-9538-y

[68] I. Kurt, M. Ture, and A. T. Kurum, "Comparing performances of logistic regression, classification and regression tree, and neural networks for predicting coronary artery disease," *Expert Systems with Applications*, vol. 34, no. 1, pp. 366–374, 2008. [Online]. Available: http://doi.org/10.1016/J.ESWA.2006.09.004

[69] "Treebagger, create bag of decision trees," (Last Accessed on 04/08/2022). [Online]. Available: https://www.mathworks.com/help/stats/treebagger.html

[70] Q. Li, F. Qiao, A. Mao, and C. McCreight, "Characterizing the importance of criminal factors affecting bus ridership using random forest ensemble algorithm," *Transportation Research Record*, vol. 2673, no. 4, pp. 864–876, 2019. [Online]. Available: http://doi.org/10.1177/0361198119837504

[71] A. Liaw and M. Wiener, "Classification and regression by randomforest," *R News*, vol. 2, no. 3, pp. 18–22, 2002. [Online]. Available: https://www.researchgate.net/publication/228451484_Classification_and_Regression_by_RandomForest

[72] "Predict," (Last Accessed on 04/08/2022). [Online]. Available: https://www.mathworks.com/help/stats/treebagger.predict.html

[73] L. Van Der Maaten and G. Hinton, "Visualizing data using t-sne," *Journal of Machine Learning Research*, vol. 9, no. 11, pp. 2579–2605, 2008. [Online]. Available: http://jmlr.org/papers/v9/vandermaaten08a.html

[74] C. Marquez de la Plata, T. Hart, F. Hammond, A. Frol, A. Hudak, C. Harper, T. O'Neil-Pirozzi, J. Whyte, M. Carlile, and R. Diaz-Arrastia, "Impact of age on long-term recovery from traumatic brain injury," *Archives of physical medicine and rehabilitation*, vol. 89, no. 5, pp. 896–903, 2008. [Online]. Available: https://doi.org/10.1016/j.apmr.2007.12.030

[75] T. Wong, "Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation," *Pattern Recognition*, vol. 48, no. 9, pp. 2839–2846, 2015. [Online]. Available: https://doi.org/10.1016/j.patcog.2015.03.009

[76] D. Kamper, A. McKenna-Cole, L. Kahn, and D. Reinkensmeyer, "Alterations in reaching after stroke and their relation to movement direction and impairment severity," *Archives of Physical Medicine and Rehabilitation*, vol. 83, no. 5, pp. 702–707, 2002. [Online]. Available: https://doi.org/10.1053/apmr.2002.32446

[77] T. J. Bovend'Eerdt, H. Dawes, H. Johansen-Berg, and D. Wade, "Evaluation of the Modified Jebsen Test of Hand Function and the University of Maryland Arm Questionnaire for Stroke," *Clinical Rehabilitation*, vol. 18, no. 2, pp. 195–202, 2004. [Online]. Available: http://doi.org/10.1191/0269215504cr722oa

[78] L. van de Ven-Stevens, M. Munneke, C. Terwee, P. Spauwen, and H. van der Linde, "Clinimetric properties of instruments to assess activities in patients with hand injury: A systematic review of the literature," *Archives of Physical Medicine and Rehabilitation*, vol. 90, no. 1, pp. 151–169, 2009. [Online]. Available: http://doi.org/10.1016/j.apmr.2008.06.024

[79] H. C. Chang, Y. L. Hsu, S. C. Yang, J. C. Lin, and Z. H. Wu, "A wearable inertial measurement system with complementary filter for gait analysis of patients with stroke or parkinson's disease," *IEEE Access*, vol. 4, pp. 8442–8453, 2016. [Online]. Available: https://doi.org/10.1109/ACCESS.2016.2633304

[80] B. Erdaş, I. Atasoy, K. Açici, and H. Oğul, "Integrating features for accelerometer-based activity recognition," *Procedia Computer Science*, vol. 98, pp. 522–527, 2016. [Online]. Available: https://doi.org/10.1016/J.PROCS.2016.09.070

Appendices

A Appendix I. IEEE/ACM CHASE 2022 International Conference Breakthrough Poster Paper

2022 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)

Wearable Technology and Machine Learning to Monitor Upper-Limb Use in Brain Injury Survivors

Cristiana Ernesto
Faculty of Sciences
University of Lisbon
Lisbon, Portugal
cristianaernesto@campus.ul.pt

Federico Parisi
Dept. PM&R
Harvard Medical School
Boston, MA, USA
fparisi@partners.org

Catherine Adans-Dester
Dept. PM&R
Harvard Medical School
Boston, MA, USA
cadans-dester@partners.org

Anne O'Brien
Dept. PM&R
Harvard Medical School
Boston, MA, USA
aobrien1@partners.org

Gloria Vergara-Diaz
Dept. PM&R
Harvard Medical School
Boston, MA, USA
gvergaradiaz@partners.org

Randie Black-Schaffer
Dept. PM&R
Harvard Medical School
Boston, MA, USA
rblackschaffer@mgh.harvard.edu

Ross Zafonte
Dept. PM&R
Harvard Medical School
Boston, MA, USA
rzafonte@mgh.harvard.edu

Hugo Ferreira
Faculty of Sciences
University of Lisbon
Lisbon, Portugal
hhferreira@fc.ul.pt

Paolo Bonato
Dept. PM&R
Harvard Medical School
Boston, MA, USA
pbonato@mgh.harvard.edu

ABSTRACT

The study explored the use of machine learning-based algorithms to estimate clinical scores (meant to capture the quality of upper-limb movements) via the analysis of wearable sensor data collected during the performance of functional tasks. Data was recorded using accelerometer units from thirty-seven individuals with upper-limb hemiparesis due to acquired brain injury while they performed eight activities of daily living. Clinicians assessed the study participants' quality of movement using the Functional Ability Scale. A hierarchical Random Forest-based algorithm was developed to derive the Functional Ability Scale scores from the accelerometer data. The results showed a high correlation (R^2=0.91) between the estimates derived from the accelerometer data and the scores provided by the clinicians. The method herein presented has the potential to provide a means to track the motor recovery process in patients experiencing upper-limb motor impairments following a stroke or a traumatic brain injury hence enabling precision rehabilitation interventions.

KEYWORDS

Acquired Brain Injury, Machine Learning, Precision Rehabilitation, Wearable Technology

ACM Reference format:

Cristiana Ernesto, Federico Parisi, Catherine Adans-Dester, Anne O'Brien, Gloria Vergara-Diaz, Randie Black-Schaffer, Ross Zafonte, Hugo Ferreira and Paolo Bonato. 2022. Wearable Technology and Machine Learning to Monitor Upper-Limb Use in Brain Injury Survivors. In *IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE), November 17-19, 2022, Washington D.C., USA.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3551455.3564713

IEEE/ACM CHASE, November, 2022, Washington, D.C., USA
 978-1-4503-9476-5/22/11...$15.00
https://doi.org/10.1145/3551455.3564713

1 Introduction

Acquired brain injury (ABI), including traumatic brain injury (TBI) and stroke, is a leading cause of long-term disability worldwide [1] and often results in severe motor impairments that negatively impact patients' independence and quality of life. The recovery of the upper-limb motor function following an ABI event is particularly important to preserve patients' ability to perform activities of daily living (ADL) and it is often the focus of rehabilitation interventions. The high variability in the response to the intervention across individuals calls for the need to develop patient-specific rehabilitation strategies. This kind of "precision rehabilitation" requires the assessment of motor gains and clinical outcomes at multiple time points to track the recovery trajectory of each patient, which can be time and resource-consuming when using traditional clinical evaluation tools. Wearable technology and machine learning (ML) enable the estimation and tracking of clinical outcome measures with minimal burden on both clinicians and patients. Previous work demonstrated the suitability of ML-based algorithms to derive clinical score estimates from wearable sensor data collected during the performance of a battery of motor task that is associated with the clinical assessment of functional limitations in stroke and TBI survivors [2]. In the study herein presented, we focused on the development of algorithms to estimate upper-limb quality of movement during the performance of simulated ADL tasks, using wearable sensor data and ML-based techniques.

Figure 1: Activities of daily living assessed in the study.

2 Methods

The study was conducted on a heterogeneous sample of 37 ABI individuals with upper-limb hemiparesis (16 stroke survivors and 21 TBI survivors). Enrolled subjects participated in two assessment visits, one at baseline (before initiating the rehabilitation program) and one at discharge (post-treatment, after at least 6 weeks). During each visit, a battery of clinical assessments was administered by a therapist. After the clinical evaluation, participants were instrumented with five wireless 3-axis accelerometers (Shimmer2r by Shimmer Sensing, Dublin, Ireland) positioned on the wrists, upper arms, and chest, and two 2-axis wired accelerometers on the thumb and index fingers. Study participants were instructed to perform a series of eight ADLs (Figure 1). A study therapist or a clinician assessed the quality of movement for each task and generated a clinical score accordingly to the Functional Ability Scale (FAS) criteria (score from 0 to 5) of the Wolf Motor Function Test (WMFT) [3].

The raw sensor data was imported in MATLAB (The MathWorks Inc, Natick, MA, USA), and processed following the steps shown in Figure 2. First, the data was processed to improve the accelerometer signal quality. This procedure included signal calibration, filtering, and segmentation. An extensive set of data features was then extracted to characterize participants' upper-limb motor behaviors. To mitigate the impact on the ML model training of the uneven distribution of the clinical scores assigned to each task, a data rebalancing technique, namely the Adaptive Synthetic sampling approach (ADASYN) [4], was used before carrying out a feature selection procedure based on the Random Forest (RF) [5] built-in feature importance method. The resulting data features and the clinical labels were then used to train in a supervised fashion a specific ML model per task. We used a hierarchical approach based on a first RF binary classifier (100 trees) to distinguish between lower (scores 1-2) and higher (scores 3-5) FAS scores, followed by a RF regression module (100 trees) to estimate a continuous value for the FAS scores between 3 and 5. A Leave-One-Subject-Out cross-validation (LOSOCV) approach was used to validate each model's performance. Lastly, the individual task predictions were aggregated using a linear model to obtain a total score as a percentage of the maximum achievable score across all tasks.

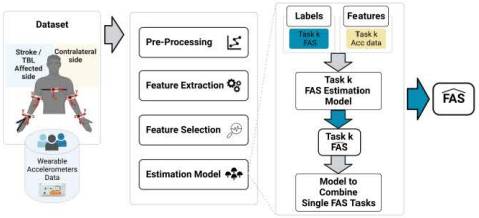


Figure 2: Processing pipeline for estimating the FAS scores.

3 Results and Discussion

The FAS scores estimated using the proposed approach showed excellent agreement with the FAS scores generated by the clinicians. Figure 3 (a) shows the linear fit between the predicted values and the clinical scores, which was marked by a coefficient of determination $R^2 = 0.91$. The analysis of bias, displayed in Figure 3 (b), showed that the estimation error falls in a range between -11% and +12% of the FAS range and that the estimation bias is negligible.

These results are encouraging as they show that accurate FAS score predictions can be derived from wearable sensor data collected during the performance of ADLs. The approach herein presented could be used to enable the design of precision rehabilitation interventions that rely on tracking the response of individual patients to the prescribed therapy and adjusting it accordingly to maximize motor gains. Furthermore, the proposed approach paves the way to transitioning from clinician-reported outcomes to sensor-based outcome estimates that are based on data collected where the impact of the intervention matters the most, namely in the home and community settings.

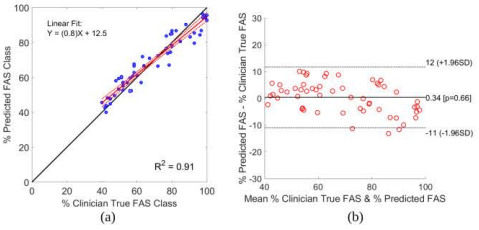


Figure 3: (a) Scatter plot of the final estimate of the FAS scores vs. the FAS scores assigned by clinicians. (b) Bland-Altman plots for the bias analysis of the estimation error.

REFERENCES

[1] V. Feigin, S. Barker, R. Krishnamurthi, A. Theadom, and N. Starkey. 2010. Epidemiology of ischaemic stroke and traumatic brain injury. In *Best practice & Research Clinical Anaesthesiology*, vol. 24, no. 4, pp. 485–494. DOI: http://doi.org/10.1016/j.bpa.2010.10.006

[2] C. Adans-Dester, N. Hankov, A. O'Brien, G. Vergara-Diaz, R. Black-Schaffer, R. Zafonte, J. Dy, S. I. Lee, and P. Bonato. 2020. Enabling precision rehabilitation interventions using wearable sensors and machine learning to track motor recovery. In *npj Digital Medicine*, vol. 3, no. 1, p. 121. DOI: http://doi.org/10.1038/s41746-020-00328-w

[3] S. Wolf, P. Catlin, M. Ellis, A. Archer, B. Morgan, and A. Piacentino. 2001. Assessing Wolf Motor Function Test as outcome measure for research in patients after stroke. In *Stroke*, vol. 32, no. 7, pp. 1635–1639. DOI: http://doi.org/10.1161/01.STR.32.7.1635

[4] H. He, Y. Bai, E. Garcia, and S. Li. 2008. Adasyn: Adaptive synthetic sampling approach for imbalanced learning. In *Hong Kong: 2008 IEEE International Joint Conference on Neural Networks*, vol. 74, no. 4, pp. 1322–1328. DOI: http://doi.org/10.1109/IJCNN.2008.4633969

[5] L. Breiman. 2001. Random forests. In *Machine Learning*, vol. 45, no. 1, pp. 5–32. DOI: http://doi.org/10.1023/A:1010933404324

B Appendix II. Functional Ability Scale

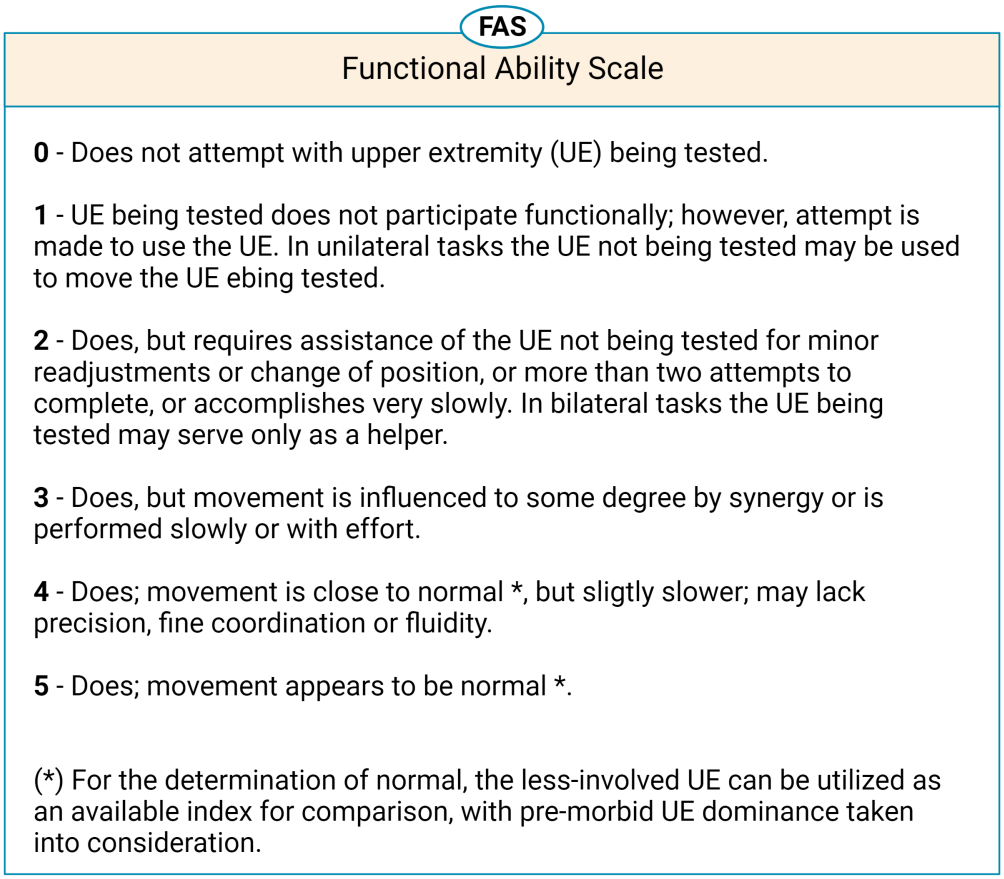


Figure B.1: Functional Ability Scale (FAS) scores descriptions. Adapted from [11].

C Appendix III. Fugl-Meyer Assessment for Upper Extremity

FMA

Fugl-Meyer Assessment for Upper Extremity

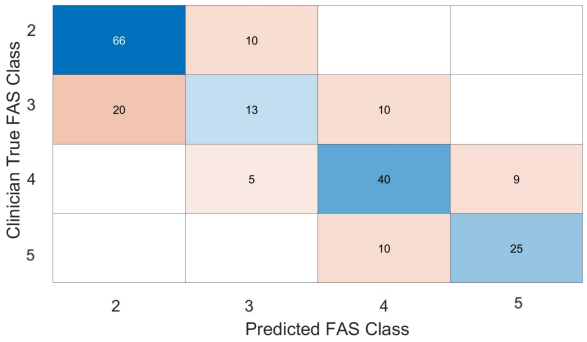
	ITEM	SCORE	SCORING
Shoulder/Elbow/Forearm	**Reflexes**		
	1. Biceps or finger flexors		0 = No reflex initiated 2 = Reflex elicited
	2. Triceps		
	Flexor synergy **(forearm supinated, bring affected arm to ipsilateral ear)**		
	3. Retraction		0 = Cannot perform at all 1 = Performed partly 2 = Performed faultlessly
	4. Elevation		
	5. Abduction ≥ 90°		
	6. External rotation		
	7. Elbow flexion		
	8. Forearm supination		
	Extensor synergy **(forearm pronated and hand to contralateral knee)**		
	9. Adduction and internal rotation		0 = Cannot perform at all 1 = Performed partly 2 = Performed faultlessly
	10. Elbow extension		
	11. Forearm pronation		
	Mixed Synergy Patterns		
	12. Hand to lumbar spine		0 = No specific action can be performed 1 = Hand must pass the ASIS 2 = Action must be done faultlessly
	13. Shoulder flexion to 90° (elbow at 0°, forearm neutral)		0 = Arm is immediately abducted or elbow flexes at start of motion 1 = Abduction or elbow flexion occurs in later phase of motion 2 = Faultless motion
	14. Forearm pronation-supination (shoulder at 0°, elbow flexed at 90°)		0 = Correct position of shoulder and elbow cannot be attained, and/or pronation or supination cannot be performed at all. 1 = Active pronation or supination can be performed even within a limited ROM and at the same time, the shoulder and elbow are correctly positioned 2 = Complete pronation and supination with correct positions at elbow and shoulder
	Isolated movements without synergy with patient sitting		
	15. Shoulder abduction to 90° (elbow at 0°, forearm pronated)		0 = Initial elbow flexion occurs or any deviation from pronated forearm occurs 1 = Motion can be performed partly, or if during motion elbow is flexed or forearm cannot be kept in pronation. 2 = Faultless motion
	16. Shoulder flexion 90°- 180° (elbow at 0°, forearm neutral)		0 = Initial flexion of elbow or shoulder abduction occurs. 1 = Elbow flexion or shoulder abduction occurs during shoulder flexion 2 = Faultless motion

Figure C.1: Fugl-Meyer Assessment for Upper Extremity scores descriptions (Part I). Adapted from [12].

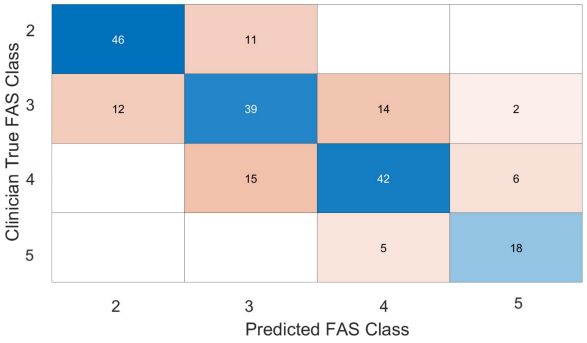
	17. Forearm pronation-supination (shoulder flexed between 30°-90°, elbow at 0°)		0 = Supination and pronation cannot be performed at all, or elbow and shoulder positions cannot be obtained. 1 = Elbow and shoulder properly positioned, and pronation and supination performed in a limited range 2 = Faultless motion
	18. Normal reflex activity of Biceps, Finger Flexors, and Triceps Score this section only if 15-18 were scored a perfect score of 6.		0 = 2/3 Reflexes are markedly hyperactive 1 = 1 Reflex markedly hyperactive or at least 2 reflexes are lively 2 = No more than one reflex is lively and none is hyperactive
Wrist (all tests, forearm pronated)	19. Wrist stability (elbow at 90°, shoulder at 0°)		0 = Pt cannot extend wrist to required 15° 1 = Extension is accomplished but no resistance is taken 2 = Position can be maintained with a light resistance
	20. Wrist flexion-extension (elbow at 90°, shoulder at 0°)		0 = Volitional movement does not occur 1 = Pt does not actively move the wrist through total ROM 2 = Faultless, smooth movement
	21. Wrist stability (elbow at 0°, shoulder flexed and/or abducted)		0 = Pt cannot extend wrist to required 15° 1 = Extension is accomplished but no resistance is taken 2 = Position can be maintained with a light resistance
	22. Alternating motions of extension and volar flexion (elbow fully extended, shoulder flexed and abducted)		0 = Volitional movement does not occur 1 = Pt does not actively move the wrist through total ROM 2 = Faultless, smooth movement
	23. Circumduction of wrist		0 = Cannot be performed 1 = Jerky motion or incomplete circumduction 2 = Complete motion with smoothness
Hand	24. Mass flexion (fist)		0 = No flexion occurs 1 = Some, but not full active finger flexion 2 = Full active flexion (compared with unaffected hand)
	25. Mass extension (from a position of full flexion)		0 = No extension occurs 1 = Patient can release an active mass flexion grasp 2 = Full active extension (compared with unaffected hand)
	26. PIP - DIP hook grasp (MP joints extended, PIPs & DIPs flexed) Grasp is tested against resistance		0 = Required position cannot be achieved 1 = Grasp is weak 2 = Grasp can be maintained against relatively great resistance
	27. Thumb adduction with paper (all other joints at 0°)		0 = Function cannot be performed 1 = Scrap of paper interposed between thumb and index finger can be kept in place, but not against a slight tug 2 = Paper is held firmly against a slight tug
	28. Pincer grasp with a pencil		0 = Function cannot be performed 1 = Object can be kept in place but not against a tug 2 = Object is held firmly against a tug
	29. Cylinder grasp with a small can		
	30. Spherical grasp with a tennis ball		
Coord./Speed	**Finger to nose 5x with eyes closed**		
	31. Tremor		0 = Marked tremor or dysmetria 1 = Slight tremor or dysmetria 2 = No tremor or dysmetria
	32. Dysmetria		
	33. Time		0 = >6 seconds difference between hands 1 = 2-5 second difference between hands 2 = < 2 second difference between hands
	TOTAL		

Figure C.2: Fugl-Meyer Assessment for Upper Extremity scores descriptions (Part II). Adapted from [12]

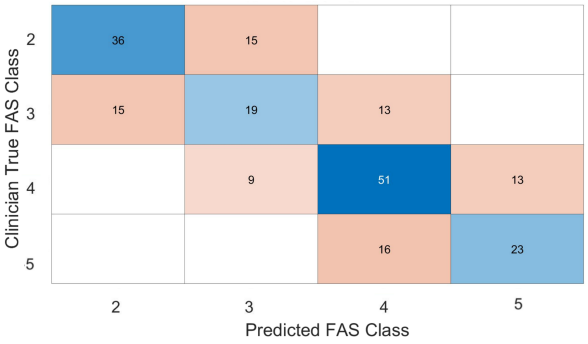
D Appendix IV. Additional Functional Ability Scale Model 3 Results



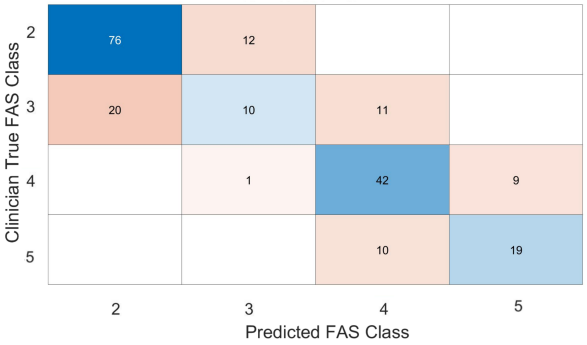
(a)



(b)

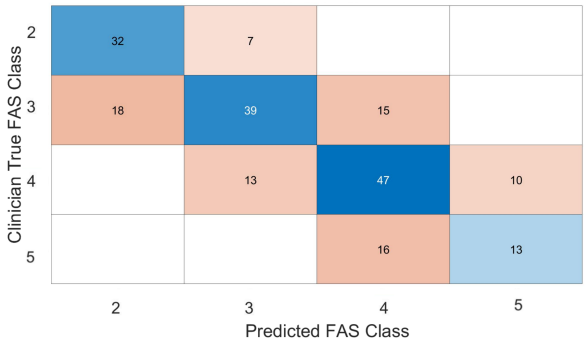


(c)

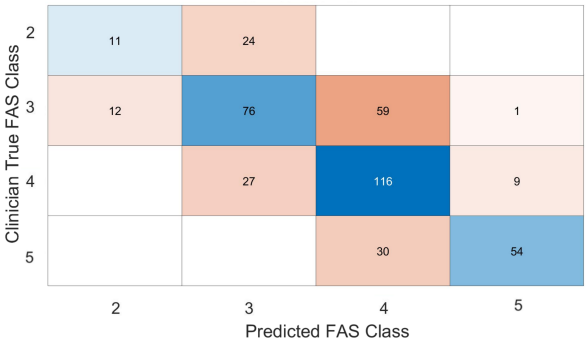


(d)

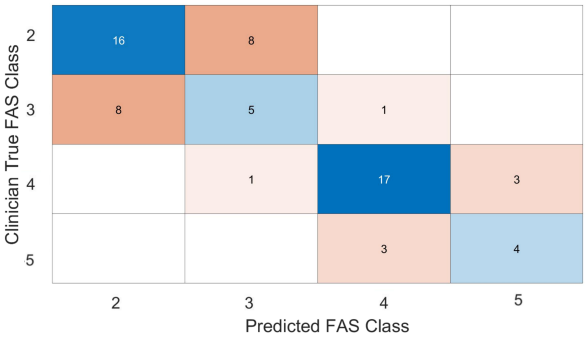
Figure D.1: Confusion Matrices concerning Model 3 Merged Hierarchical results to predict Tasks 1-4 Single Task Scores concerning FAS classes 2 to 5. (a) Task 1. (b) Task 2. (c) Task 3. (d) Task 4.



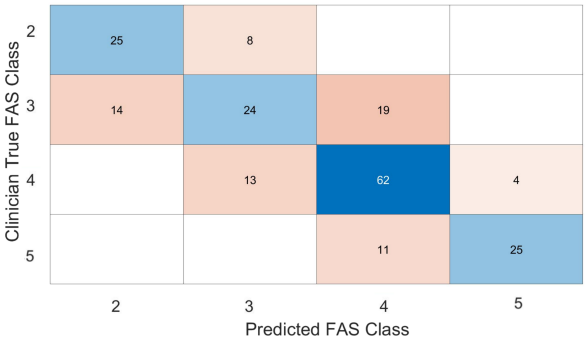
(e)



(f)



(g)



(h)

Figure D.2: Confusion Matrices concerning Model 3 Merged Hierarchical results to predict Tasks 5-8 Single Task Scores concerning FAS classes 2 to 5. (e) Task 5. (f) Task 6. (g) Task 7. (h) Task 8.

E Appendix V. Additional Fugl-Meyer Assessment Model 1 Results

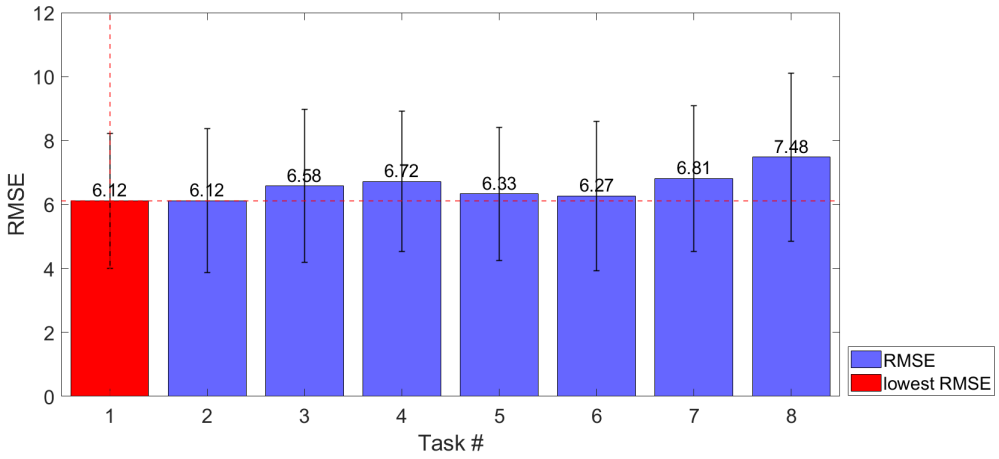


Figure E.1: RMSE of the FMA Single Tasks Scores estimated with FMA Model 1. Emphasized in red is the lowest RMSE taking into account all Tasks comparisons.

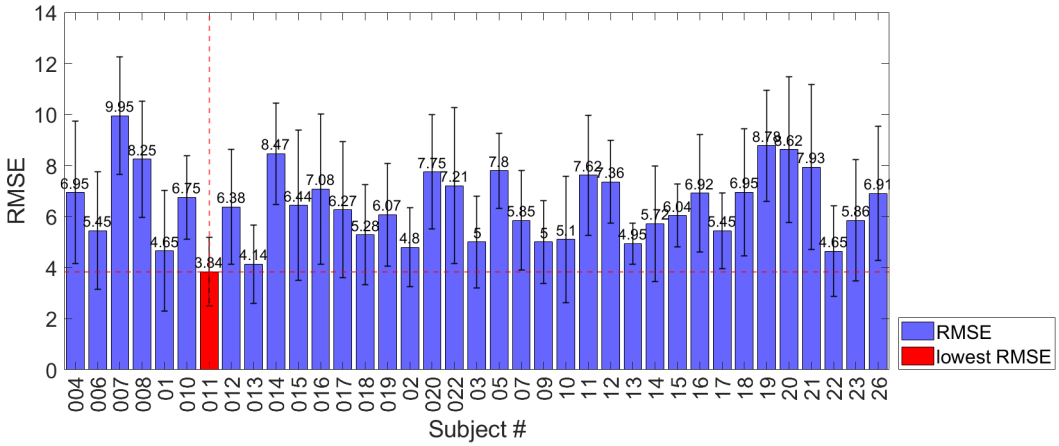


Figure E.2: RMSE of the FMA Single Tasks Scores estimated with FMA Model 1 for each patient. Emphasized in red is the lowest RMSE taking into account all subjects comparisons.