Imperial College London

EPSRC Centre for Doctoral Training in NEUROTECHNOLOGY for life and health

Wearable Fusion System for Assessment of Motor Function in Lesion-Symptom Mapping Studies

Lewis Formstone

Department of Mechanical Engineering
Imperial College of Science, Technology and Medicine
London, UK

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

This thesis was supervised by
Dr. Ravi Vaidyanathan, Department of Mechanical Engineering

and co-supervised by
Dr. Paul Bentley, Department of Brain Sciences
Prof. Etienne Burdet, Department of Bioengineering
Prof. Alison McGregor, Department of Surgery & Cancer

PUBLISHED BY THE IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE

First Edition, September 2020

Statement of Originality

I, Lewis Formstone, declare that all of the work presented in this thesis is my own with the exception of any parts which are referenced according to the original author.

Acknowledgements

Special thanks are given to all the participants at Charing Cross Hospital who agreed to take part in this study. The resilience they showed, and the patience extended to me during a stage of such upheaval in their lives is greatly appreciated and enabled this study to be possible. I would also like to give thanks to all my supervisors who all provided critical insights and helped keep me on the right path. Finally, thanks to all my family and friends for their support which helped maintain my sanity in the more difficult stages. Special mention goes to my mother who managed to achieve every qualification bar a PhD, hopefully she can achieve satisfaction from this submission vicariously.

Abstract

Lesion-symptom mapping studies are a critical component of addressing the relationship between brain and behaviour. Recent developments have yielded significant improvements in the imaging and detection of lesion profiles, but the quantification of motor outcomes is still largely performed by subjective and low-resolution standard clinical rating scales. This mismatch means than lesion-symptom mapping studies are limited in scope by scores which lack the necessary accuracy to fully quantify the subcomponents of motor function.

The first study conducted aimed to develop a new automated system of motor function which addressed the limitations inherent in the clinical rating scales. A wearable fusion system was designed that included the attachment of inertial sensors to record the kinematics of upper extremity. This was combined with the novel application of mechanomyographic sensors in this field, to enable the quantification of hand/wrist function. Novel outputs were developed for this system which aimed to combine the validity of the clinical rating scales with the high accuracy of measurements possible with a wearable sensor system. This was achieved by the development of a sophisticated classification model which was trained on series of kinematic and myographic measures to classify the clinical rating scale. These classified scores were combined with a series of fine-grained clinical features derived from higher-order sensor metrics.

The developed automated system graded the upper-extremity tasks of the Fugl-Meyer Assessment with a mean accuracy of 75% for gross motor tasks and 66% for the wrist/hand tasks. This accuracy increased to 85% and 74% when distinguishing between healthy and impaired function for each of these tasks. Several clinical features were computed to describe the subcomponents of upper extremity motor function. This fine-grained clinical feature set offers a novel means to complement the low resolution but well-validated standardised clinical rating scales.

A second study was performed to utilise the fine-grained clinical feature set calculated in the previous study in a large-scale region-of-interest lesion-symptom mapping study. Statistically significant regions of motor dysfunction were found in the corticospinal tract and the internal capsule, which are consistent with other motor-based lesion-symptom mapping studies. In addition, the cortico-ponto-cerebellar tract was found to be statistically significant when testing with a clinical feature of hand/wrist motor function. This is a novel finding, potentially due to prior studies being limited to quantifying this subcomponent of motor function using standard clinical rating scales. These results indicate the validity and potential of the clinical feature set to provide a more detailed picture of motor dysfunction in lesion-symptom mapping studies.

Contents

List of Figures

Nomenclature

AC Anterior Commissure

AdaBoost Adaptive Boosting

ADC Analogue-to-Digital Converter

AICHA Atlas of Intrinsic Connectivity of Homotopic Areas

ALI Automated Lesion Identification

ALL Automated Anatomical Labelling

ANN Artificial Neural Network

ARAT Action Research Arm Test

CI Clustering Index

CPCT Cortico-ponto-cerebellar Tract

CRST Corticoreticulospinal Tract

CSF Cerebrospinal Fluid

CST Corticospinal Tract

CSV Comma-separated Values

CT Computed Tomography

DOF Degrees of Freedom

DWI Diffusion-Weighted Imaging

EMG Electromyogram

FLAIR FLuid Attenuated Inversion Recovery

FMA Fugl-Meyer Assessment

FMA-UE Upper extremity section of the Fugl-Meyer Assessment

fMRI functional Magnetic Resonance Imaging

FSR Force Sensing Resistor

GCS Glasgow Coma Scale

GUI Graphical User Interface

ICBM International Consortium of Brain Mapping

iEMG Intramuscular Electromyogram

IIR Infinite Impulse Response

IMU Inertial Measurement Units

KNN K-Nearest Neighbours

Lasso Least absolute shrinkage and selection operator

LDA Linear Discriminant Analysis

lesionGnb Gaussian naïve Bayes lesion detection method

LightGBM Light Gradient Boosting

LINDA Lesion Identification with Neighbourhood Data Analysis

LSM Lesion-Symptom Mapping

MAS Modified Ashworth Scale

MIPS Million Instructions per Second

MMG Mechanomyogram / Mechanomyography

MNI Montreal Neurological Institute

MPF Mean Power Frequencies

MRI Magnetic Resonance Imaging

MVC Max Voluntary Contraction

NPM Non-parametric Mapping Software

PCA Principal Component Analysis

PCB Printed Circuit Board

RLSM Region-based Lesion-Symptom Mapping

RMS Root Mean Square

SDK Software Development Kit

sEMG Surface Electromyogram

SIS Stroke Impact Scale

SLA Stereolithography

SM Spectral Moments

SPM Statistical Parametric Mapping

SVM Support Vector Machine

TIA Transient Ischaemic Attack

UART Universal Asynchronous Receiver Transmitter

UID Unique Identifier

VBM Voxel-based Morphometry

VLSM Voxel-based Lesion-Symptom Mapping

WMFT Wolf Motor Function Test

CHAPTER

ONE

Introduction

1.1 Motivation

Over the past few decades stroke has increasingly become one of the major causes of acquired adult disability in the UK [22]. This rate of increase is primarily due to improvements in the quality of post-stroke care which has increased the proportion of individuals surviving stroke and living with a resulting disability. This increase in quality is driven by several recent innovations including imaging technology, thrombolysis, and surgery techniques. Despite these innovations, methods of quantifying the motor deficits post-stroke remain relatively unchanged in decades and are still dependent on subjective evaluation by a clinician. The key drawbacks associated with this method include high time expenditure, the requirement for extensive training, and low resolution. These limitations have a significant impact in terms of the quality of monitoring and targetting patient rehabilitation. Secondly, the reliance on these limited clinical rating scales has ramifications on the depth of insights possible in motor-based research studies. This includes lesion-symptom mapping studies which seek to study the relationship between brain lesion profiles and measures of motor dysfunction but are limited by the low-resolution scores provided by clinical rating scales.

Multiple systems have been trialled over the past decade to attempt to automate the traditional clinical rating scales. These systems have attempted to improve on the limitations of these rating scales in one or a combination of accuracy, time saving, and objectivity. Developed systems have taken the form of sensors which collect data from the subjects and algorithms which are used to compute motor scores. A pervasive limitation

of prior automated systems is the difficulty faced in quantifying hand and wrist function. This is a major component of stroke dysfunction but previously proposed methods such as the use of instrumented gloves or visual-sensing suffer from limitations in practicality or hygiene. Other limitations of prior systems which have prevented their clinical uptake include lack of useful information (beyond predicting a clinical score), extensive time required for setup, and the narrow range of motor dysfunction levels which may be tested.

Lesion-symptom mapping studies form a key tool in improving the understanding of functional neuroanatomy of the brain. These studies have been performed in the past to improve understanding of a number of cognitive processes including spoken language, somatosensory function, and motor function. The insights possible via lesion-symptom mapping have greatly increased in recent years due to huge leaps in our ability to image the brain in-vivo. Despite these advances, measures of motor function are still dependent on clinical rating scales which have remained unchanged for many years. Since the insights possible via lesion symptom mapping are directly dependent on the quality of the outcome measures provided; these clinical rating scales presently form one of the major limitations of these studies.

Sensor systems, such as those proposed to supplement or replace the clinical rating scales, offer the ability to calculate metrics of motor function which would either be impossible or prohibitively difficult for a clinician to calculate. For instance, whilst a clinician may be able to assign an approximate rating for how well a subject may be able to flex their shoulder, it is possible for a sensor system to instead assign this with a higher resolution value in terms of degrees rotation. In addition, sensor systems could be used to record metrics that a clinician cannot freely observe, such as the use of myographic sensors to measure subject muscle activity. It is expected that novel and higher quality motor outcome measures which a sensor-based system would offer a whole new domain of potential insights into lesion-symptom mapping studies. Prior studies which have already implemented sensors have been limited to simple setups such as the use of a hand dynamometer. There is clearly room for significant innovation in this field by the use of more complex derived motor metrics from a sophisticated sensor system.

1.2 Research Focus

The first study included in this thesis aimed to build on prior systems towards the development of an automated system for scoring motor function post-stroke. There are several key drawbacks of previous works in this field which should be addressed in the present study to enable clinical validity. The first drawback is the low statistical power present in the majority of previous studies due to small sample sizes of fewer than 20 subjects. Secondly, poor quantification of arm motion was present in studies which considered data collected from a single inertial measurement unit attached to the wrist to be representative of the entire arm motion. Finally, many prior studies have either ignored or poorly quantified the grasp and hand function of the subject, a major paradigm of post-stroke weakness. These drawbacks were addressed in the current study by the use of a large subject pool, full instrumentation of the upper arm, and the use of myographic sensing respectively. A secondary component of the study was the computation of a series of fine-grained clinically relevant features which provided additional insight into the subcomponents of motor function.

The second study presented in this thesis involved the incorporation of the wealth of information gathered by the sensors proposed in the prior study to perform a lesion-symptom mapping study. The majority of earlier motor-based lesion-symptom mapping studies have utilised either clinical rating scales or simplistic sensors. This would be the first study to utilise motor outcome measures derived from the fusion of sensor data collected across the upper extremity. Motor outcome measures were selected from the collection of orientation and myographic features calculated from the different body segments in the previous study. Features were selected based on a pre-prescribed criteria including evidence of usefulness as metrics of motor function. Lesions were identified and boundaries drawn onto subject CT and MRI scans. Finally, a large-scale lesion-symptom mapping study was conducted by applying these novel motor outcome measures derived from the sensors in addition to the FMA clinical rating scale, which was used as a baseline measure. Significant results were considered for regions which were identified as being statistically significant for the novel motor outcome measures but not for the baseline FMA. This result would evidence the additional information provided by sensor derived motor outcome measures than by the FMA alone.

1.3 Thesis Structure

This thesis is structured into five main chapters as follows:

- *Chapter 2* details all the background information required to fully appreciate the work conducted in the two studies presented in this thesis.
- *Chapter 3* follows the development of an automated system for rating upper arm motor function post-stroke. A literature review is provided to supply details of prior automated systems developed in this domain. This is followed by the system development stages and a large clinical study conducted to validate the system. The motor outcomes produced by the system, composed of a classified FMA-UE score and a series of fine grained features, will be discussed in depth.
- *Chapter 4* details a large motor-based lesion-symptom mapping study utilising the fine grained features developed in the previous chapter as the motor outcome measures. A literature review is included which outlines the techniques developed in prior lesion-symptom mapping studies and their major findings. Next, the major pieces of software used for processing and analysis within the study are discussed. Finally, the lesion symptom mapping results are discussed with respect to the findings using motor outcomes provided by the fine grained features (developed in the previous chapter) as well as the overall FMA-UE (as a baseline measure).
- *Chapter 5* presents the conclusion of the thesis. This is structured in terms of the performance of the automated system and how useful the motor outcomes provided by the fine grained features translated into the large lesion-symptom mapping study. Additionally, the novel application of mechanomyographic sensing in a wearable system for targetting motor function in stroke will be discussed.
- *Chapter 6* provides context of the findings of two major studies outlined in this thesis in terms of the wider literature and real-world applications. Any limitations of the both the wearable system and the lesion symptom mapping study are discussed, and suggestions made on how these could be improved upon in the future. Finally, the author's opinion on the future direction of such wearable systems will be summarised.

CHAPTER

TWO

Background

2.1 Chapter Structure

This chapter covers the required background topics for each of the two major studies included in this thesis. The novel automated system developed in this paper may be better understood by reading the included sections on clinical evaluation of stroke, diagnostic myography, motion tracking, defining orientation, and the use of machine learning in healthcare. The fundamentals of lesion-symptom mapping are explained via sections on the pathophysiology of stroke, neuroimaging of stroke, neuroanatomy, and the background of lesion-symptom mapping.

2.2 Stroke

There are on average over 100,000 incidences of stroke in the UK every year. Improvements in healthcare response of stroke has steadily improved leading to a present survival rate of over 80% in England, Wales, and Northern Ireland. The has resulted in a large post-stroke population in the UK of over 1.2 million people. Post-stroke care is a key component for this population with a third of stroke survivors experiencing depression post-stroke and two thirds leaving hospital with a disability. The total cost of stroke to society in the UK is estimated to be around £26 billion per year [22].

2.2.1 Pathophysiology of Stroke

Stroke is a form of cardiovascular disease which impairs the blood supply to the brain resulting in regions of brain cell death (lesions). There are several major symptoms which indicate stroke including deficits in motor function, sensation, cognition, and communication. The two types of stroke which may occur are ischaemic and haemorrhagic. The ischaemic type of stroke is most common and is caused by a lack of blood flow to the brain due to the formation of blood clots. A temporary blockage of blood flow which does not result in permanent damage is referred to as a Transient Ischaemic Attack (TIA). The Haemorrhagic type of stroke is less common and is caused by seepage of blood through a hole in a blood vessel wall into the brain or space around the brain [23]. Both forms of stroke are illustrated in Figure 2.1.

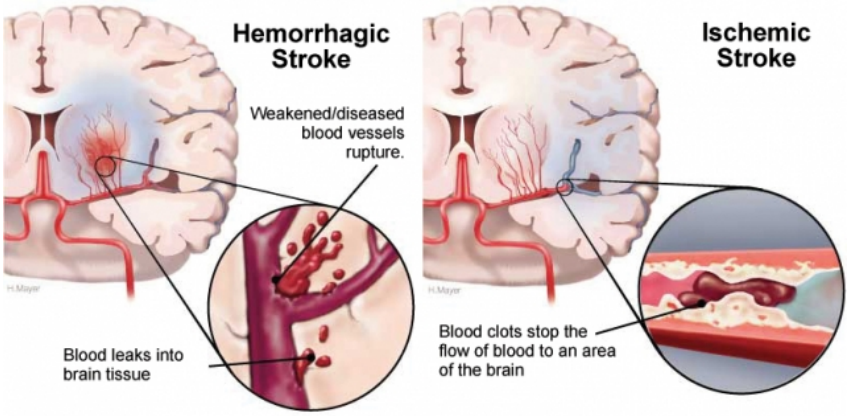


Figure 2.1: Pathophysiology of Haemorrhagic and Ischaemic Stroke Types. Taken from [1]

2.2.2 Sensorimotor Symptoms of Stroke

Weakness or paralysis is the predominant impairment that contributes to motor dysfunction post-stroke [24]. This is a consequence of disturbed signal transmission from the motor cortex to the spinal cord required for the execution of movements. The result of this is delayed initiation and termination of muscle contraction [25] and a slowness in developing forces [26]. Weakness may be present in all muscle groups of the upper limb or may selectively affect some more than others. The symmetrical nature of the motor pathways in the brain (whereby motor regions of the brain control body segments on the opposite side) means that it is common for motor dysfunction to be more pronounced on

one side, in a condition known as hemiplegia.

Sensory loss is common post-stroke across tactile, proprioceptive and high-order sensory modalities such as vision. Sensory loss after stroke is commonly associated with the degree of motor weakness but may also occur in isolation due to lesion in specific brain regions (such as the parietal cortex). Sensory loss may also have a knock-on effect on motor function. This is because the planning and execution of voluntary movement requires information regarding body positions and the prediction of future positions. This can only be achieved by integrating a variety of sensory inputs with ongoing and planning motor activity. Subjects who have lost one or more of their senses may show profoundly affected motor function even if their motor strength remains unaffected [27].

Another common motor impairment suffered post-stroke is spasticity. This motor disorder is characterised by a velocity-dependent increase in muscle tone with exaggerated tendon jerks, resulting from hyperexcitability of the stretch reflex. This impairment becomes more common in relation to the time since stroke onset and is a secondary effect of weakness and mobility on skeletal muscles [28]. Subjects with spasticity exhibit impaired functions and decreased quality of life. Another possible consequence is abnormal postural patterns due to imbalance in agonist and antagonist strength [29].

A final common impairment present post-stroke are abnormal motor synergies. This describes functional impairment due to abnormal descending motor commands post-stroke [30]. This impairment is characterised by a difficulty in coordination of different muscle groups and has been found to be independent of motor weaknesses.

2.2.3 Clinical Evaluation of Upper Extremity Motor Function

Clinical evaluations of motor function post-stroke form a critical component of optimising rehabilitation programs, managing patient expectations, and assessing outcome measures for clinical studies. Upper extremity motor function assessments may be assessed at the bedside and so are often performed as an initial method of gauging motor function immediately post-stroke. This contrasts with lower extremity assessment which requires subjects to be able to get out of bed and as such may not always be assessed in the acute stage of stroke.

There are currently more than 53 pre-existing clinical measures of upper extremity motor function currently available [31]. These may be broadly placed into three categories [32]. Firstly, scales which measure impairments to bodily function or structure. A common example of this is the Modified Ashworth Scale (MAS) which is a measure of

spasticity. Secondly, there are scales which measure dysfunctions or limitations in performing a given activity. An example of this is the Action Research Arm Test (ARAT), an upper extremity scale which assesses tasks in the categories of grasp, grip, pinch, and gross movement. Another example is the Wolf Motor Function Test (WMFT). The scale assesses upper extremity function via 21 tasks in three categories, time, functional ability, and strength. Tasks are assessed using an ordinal scale which ranges from 0-5. Finally, there are scales which define a limitation in general life. An example of this is the Stroke Impact Scale (SIS), a self-reported questionnaire which evaluates disability and health-related quality of life after stroke.

One of the most widely adopted clinical scales of impairments to bodily function post-stroke is the Fugl-Meyer Assessment (FMA) [33]. The FMA is also one of the most in-depth and well-validated rating scales of sensorimotor function post-stroke. The FMA contains tasks which covers the function of the entire body, but a shortened version is often used which only includes upper extremity motor function assessment (FMA-UE). This section examines each component of the upper extremity in isolation as well as combined in synergistic and non-synergistic movements. In addition, reflexes and coordination are recorded. Each task may be scored as a single component or the combination of multiple subcomponents depending on its complexity [33]. Each component is given a qualitative rating of either 0, 1, or 2 depending on how well it was performed. A score of 0 represents no movement during the component, a score of 1 indicating partial completion of the component, and a score of 2 denotes the component to have been performed fully.

The tasks of the FMA-UE may be split into two classes. The first of these involve gross upper arm motor function involving both the upper and lower arm segments as well as the torso in some cases. The second is the hand/wrist tasks which predominantly activate those two regions only. An illustration of how these tasks have been separated for the present study is shown in Table 2.1.

2.3 Neuroimaging of Stroke

The development of methods to image the brain in vivo has enabled a revolution in post-stroke diagnosis and care. Two imaging modalities, Computed Tomography (CT) and Magnetic Resonance Imaging (MRI), have found widespread use in this domain due to their ease of application, resolution, and sensitivity to tissues pertinent to stroke mechanisms. These two modalities will be discussed in this section with respect to their different

Table 2.1: Tasks from the FMA-UE clinical rating scales categorised as gross and hand/wrist tasks. Hand/wrists tasks are further segmented into those which predominantly involve the wrist (red background) and hand (blue background)

Gross motor tasks	Hand/wrist motor tasks
Flexor synergy	Pronation-supination (elbow at 90)
Extensor synergy	Pronation-supination (elbow at 0)
Hand to lumbar spine	Stability at 15 dorsiflexion (elbow at 90)
Shoulder flexion (0-90)	Repeated dorsi-volar flexion (elbow at 90)
Shoulder abduction	Stability at 15 dorsiflexion (elbow at 0)
Shoulder flexion (90-180)	Repeated dorsi-volar flexion (elbow at 0)
Coordination-speed	Circumduction
	Mass flexion
	Mass extension
	Hook grasp
	Thumb adduction
	Pincer grasp
	Cylinder grasp
	Spherical grasp

imaging types and their relative advantages for characterising stroke. In addition, the optimal time for imaging stroke will be discussed. This is of particular importance for research applications since different imaging modalities may over or under-estimate lesions based on the stroke stage.

2.3.1 Computed Tomography

CT has historically been and remains the most frequently used imaging modality for immediate assessment after suspected stroke. CT has the advantage of being readily available and fast to administer so may be utilised rapidly to assess the early signs of stroke. In addition, CT is used to exclude the possibility of or to assess the progression of haemorrhagic stroke due to its high sensitivity to blood. CT may be also administered with a contrast agent, and as such is known as contrast CT. This may be used to provide clearer images to identify occluded blood vessels (CT angiogram) or to detect the amount and speed of blood flow so that salvageable areas of the brain can be detected (CT perfusion) [34].

2.3.2 Magnetic Resonance Imaging

The advent of MRI has further improved the potential resolution of brain imaging. MRI has historical limitations compared to CT, including long assessment times and lack of availability but these have improved over time making it suitable for most patient populations. Observed advantages of MRI over CT include the use of non-ionising radiation and the potential for higher resolution imaging. Diffusion-Weighted Imaging (DWI) techniques, which use the diffusion of water molecules to achieve high image contrast, are becoming the gold-standard for acute lesion imaging. This is due to this modality possessing a sensitivity and specificity to acute stroke lesion which is superior to CT [2]. Another common MRI technique is the use of the FLuid Attenuated Inversion Recovery (FLAIR) sequence. This sequence removes the cerebrospinal fluid (CSF) from the resulting images and has application in imaging a variety of conditions including stroke. An example of the increased sensitivity to acute lesion which is possible with MRI as compared to CT is shown in Figure 2.2.

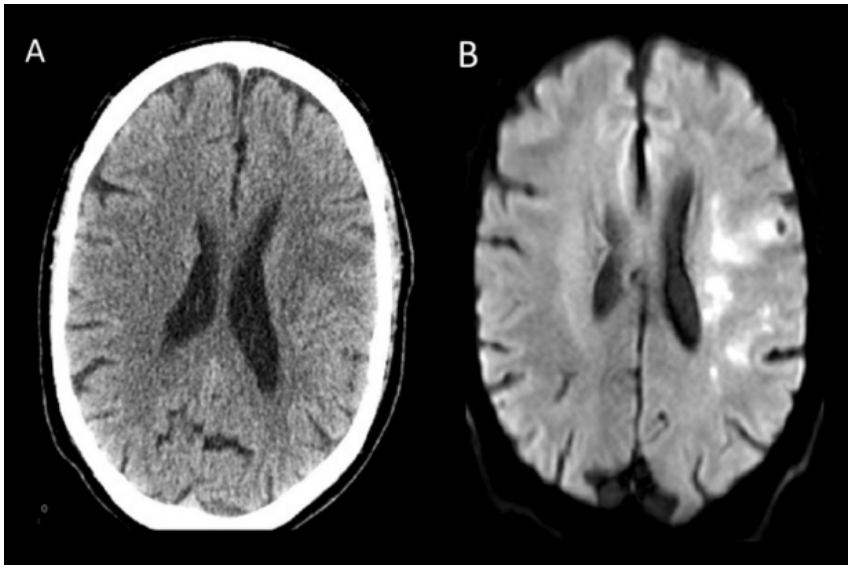


Figure 2.2: Example of CT (left) and DWI MRI (taken afterwards on the same day, right) taken from the same subject. Large lesion evident on the MRI scan as compared to slight hypodensity visible on the CT. Taken from [2]

2.3.3 Time Point of Neuroimaging

Clinical imaging of stroke is typically performed as early as possible in the acute stage to identify the severity and to exclude factors such as haemorrhage or TIA. When imaging for research purposes, there is debate about the best time to image. The reason for this is due to changes in the brain which occur after stroke. During the acute phase, brain structures are largely unmodified, but this will change in the chronic phase due to tissue reabsorption which causes structural distortions, sulcal widening, and widening of the ventricle [35]. One limitation of many prior neuroimaging studies is a failure to disclose when imaging was administered, or a definition of what time periods relate to which stroke stage. This is rectified in this thesis by the use of a set of standardised time periods as selected from the work by Allen et al [36] and presented in Table 2.2. It is worth noting here that these stages of stroke have been selected on the basis of presenting defined changes in brain structure as observed via neuroimaging. For behavioural studies, changes in outcomes generally lag observed changes in brain structure and therefore stroke stages are defined over significantly different time periods.

Table 2.2: Neuroimaging temporal stages of stroke. Taken from Allen et al [36]

Stroke stage	Time period
Hyper-acute	0 - 24 hours
Acute	24 hours - 1 week
Sub-acute	1 - 3 weeks
Chronic	$>$ 3 weeks

2.3.4 Time Point of Examination

Another significant factor when conducting a neuroimaging study, and one which varies markedly between studies [35], is the time point at which behavioural outcome measures are assessed. This can have significant ramifications in a study since a subject may present very different behavioural outcomes in the acute as opposed to the chronic phase post stroke, due to a combination of rehabilitation interventions and the natural plasticity of the brain. For this thesis subjects stroke stage when assessed was defined using the same time periods as defined for neuroimaging stage as shown in Table 2.2.

2.4 Neuroanatomy

A fundamental understanding of neuroanatomy is required to better appreciate the findings of the lesion-symptom mapping study included in this thesis. This includes an overview of the brain, including the subregions which are primarily responsible for the generation of signals related to motor function. Secondly the major motor pathways that help transmit these signals to the spinal cord, and the brain regions that these pathways travel through will be discussed.

2.4.1 Anatomical Overview

The majority of the brain may be defined in terms three large anatomical regions. The uppermost section of the brain is known as the cerebrum. This the largest part of the brain and has a role in a variety of high-level functions including memory, thoughts, the initiating of voluntary motor activities. The cerebellum lies inferior to the cerebrum. This region has a variety of functions, many of which are still poorly understood. This cerebellum is however known to have a significant role in motor and posture control. Finally, there is the brainstem consisting of the subregions: midbrain, pons, and medulla. The brain stem is understood to have a role in relaying messages to the spinal cord from the other brain regions as well as several autonomic functions [37].

The cortex (outer surface) of the cerebrum may be further segmented into different anatomical lobes. The lobes are understood to approximately correspond to different base functions. An illustration of the different lobes of the cerebral cortex as well as the cerebellum and brain stem is shown in Figure 2.3. The frontal lobe is known to be responsible for a number of cognitive skills such as memory, planning, and language. The parietal lobe is primarily responsible for receiving and processing sensory input. The somatosensory cortex is a subset of the parietal lobe and is the main receptive area for the sense of touch. The temporal lobe has a role in the derivation of visual memory, language comprehension and emotion association. Finally, the primary role of the olfactory bulb and occipital lobe are in smell and vision respectively [37].

2.4.2 Motor Cortices

Two important regions of the cortex for motor function are the primary and secondary motor cortices. These regions combine to generate signals required for the planning and

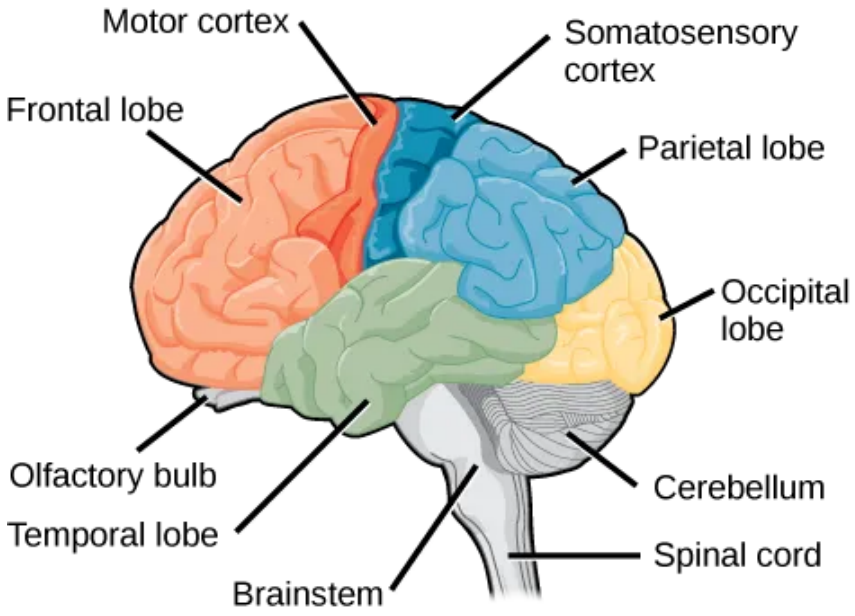


Figure 2.3: Basic anatomy of the brain, including brain lobes. Taken from [3]

execution of movements. The primary motor cortex is a subset of the frontal lobe. The primary motor cortex at each side of the brain contains a motor representation of each body part of the opposite side of the brain. The secondary cortex is composed of the premotor cortex (also located in the frontal lobe), posterior parietal cortex, and the supplementary motor area. The function of the premotor cortex is not well understood but is believed to have a role in control of the trunk. The posterior parietal cortex plays a role in planned movements and spatial reasoning. Finally, the supplementary motor area is understood to have a role in postural stabilisation and coordination of bi-manual movement [37].

2.4.3 Motor Pathways

Motor function may be defined by the neural pathways which enable motor signals to be generated and delivered to the muscles of interest. In this subsection several such pathways which play an important role in normal motor function will be described.

One neural pathway which has been identified as having an important role is the corticospinal tract (CST). The genus of this pathway is the primary and secondary motor cortices of the cerebral cortex. Motor neurons which originate at these locations then pass through the posterior limb of the internal capsule. Next the neurons pass the midbrain, and then the brain stems (pons and medulla). Finally, the tract reaches the spinal cord

where neurons synapse directly into alpha motor neurons for muscle control. The primary role of the CST is understood to be control of the distal extremities (including fine motor control of the hands) [38]. A diagram of the pathway of this tract is shown in Figure 2.4.

Another pathway which has been identified as having an important role in motor function is the Corticoreticulospinal Tract (CRST). The CRST is made up of the corticoreticular pathway and the reticulospinal tract. The origin of the corticoreticular fibres is the premotor cortex and supplementary motor area. The fibres descend along with CST and terminate in the pons. The CRST innervates the proximal extremities and axial muscles and hence has a role in postural control and locomotor function [38].

Both the CST and the CRST largely follow a common pathway in the brain and play a critical role in motor function. This means that damage to this pathway rarely affects only one of these tracts in isolation. As a result lesions which occur to this pathway may result in a deficit to both the distal and proximal extremities. For this reason, the effect of lesion load of these pathways on both short- and long-term motor outcomes has been studied extensively [39, 40, 41].

A final pathway worth mentioning due to its relevance in this thesis is the cortico-ponto-cerebellar tract (CPCT). This pathway originates in the cerebral cortex and then descends through the corona radiata and internal capsule before terminating in the pons. The CPCT is understood to be involved in the communication between the cerebellum and prefrontal cortex for the coordination and planning of motor tasks.

2.5 Lesion Symptom Mapping

Lesion-Symptom Mapping (LSM) studies are those which seek to form a relationship between the location or profile of a brain lesion and a given outcome measure. CT or MRI scans are demarcated to extract lesion maps which define the location and size of the lesions. Lesion maps are then normalised to a standard brain template. Finally, statistical analyses are used to detect statistically significant relationships between the lesion profiles and outcome measures. In this way insights into the functioning of the brain may be uncovered for the betterment of scientific knowledge and patient rehabilitation.

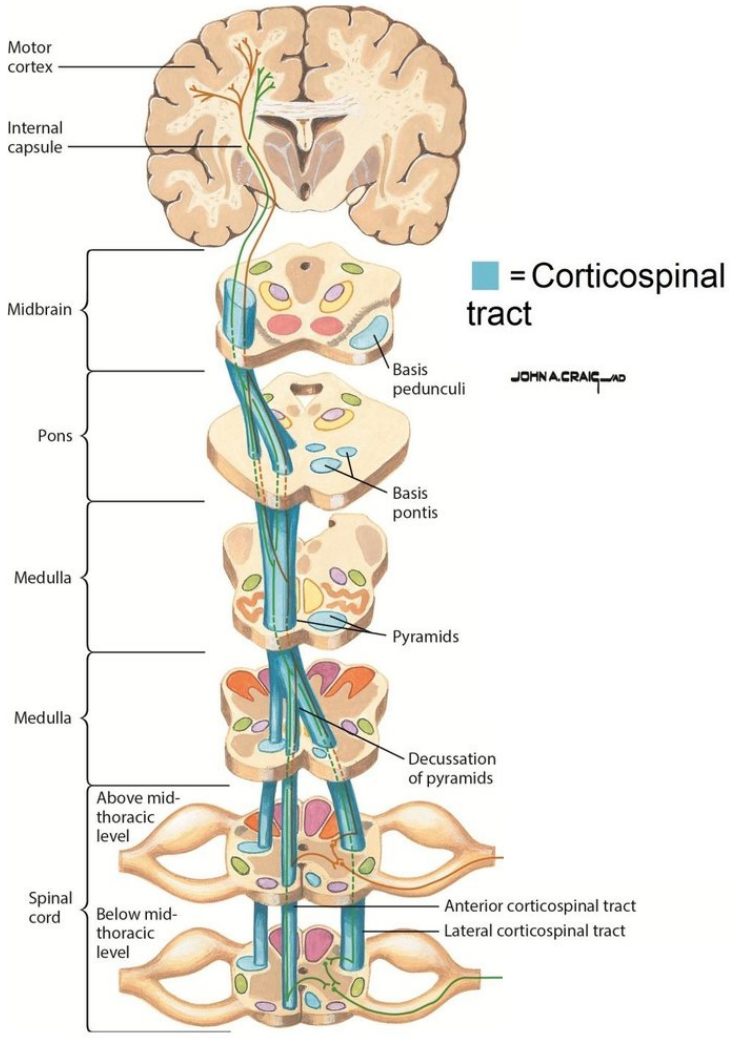


Figure 2.4: Corticospinal tract (blue) conveying motor signals from motor cortex to skeletal muscles. Adapted from [4]

2.5.1 Lesion Demarcation

The clinical evaluation of stroke for treatment typically involves only visual assessment of the lesions present in the MRI scan slices. For research applications, it is necessary to be able to demarcate these lesions in a standardised way which would enable comparisons to be made between subjects. The traditional, and gold-standard, method for demarcation is manual tracing of the regions of interest. This method requires a trained clinician to manually identify healthy/ diseased tissue and segment these using suitable software. A drawback of the manual method is that it is highly time consuming since each slice of the scan must be demarked. More recently, automated methods of demarcation have been developed which offer significant time savings. One way this has been implemented is by using a predefined template which is superimposed onto a brain region to measure the volume of healthy/ diseased tissue within the template [42]. Although automated

methods have great potential in this field, currently they do not offer comparable levels of accuracy to manual methods and may suffer from false detection due to imaging artefacts.

2.5.2 Normalisation

Normalisation in this context refers to the standardisation of either the scan slices or the demarked lesion map (if lesion maps have already been drawn at this stage). Normalisation involves the transformation from the native space (coordination frame of the scanner at data acquisition) to a stereotaxic 3D coordinate frame. This provides a common space from where comparisons or mapping studies may be made from a variety of different subjects and imaging modalities. The two requirements for image normalisation are a pre-defined 3D stereotaxic space, and a mapping function to transform the image into this space. The stereotaxic space is provided by a template or atlas which defines the target for the transformation. The mapping function is generated using suitable algorithms and defines the transformation between the native and stereotaxic space [43].

2.5.3 Analysis Types

Several methods of analysing the relationship between lesions and outcome measures have been developed. These methods vary significantly in complexity, ranging from simple overlay of the lesion maps to statistical analyses and machine learning approaches which are performed on the voxel level.

Overlay and Subtraction

This form of analysis involves the overlay of subtractions of lesion maps from different subject populations to discover insights about some function. Subjects are grouped by a certain outcome measure (for instance possession or non-possession a particular behavioural deficit). For overlay analysis, all the lesion maps of the deficit group would be overlaid and the same would be performed for the non-deficit group. A comparison would then be made between the common areas of infarctions between these two groups to identify where damage to a particular region corresponds to the behavioural deficit. For subtraction analysis, the lesion maps for the non-deficit group are instead subtracted from the deficit group. This results in a lesion map overlay which would only display lesions present in the deficit group. A limitation of overlay methods is that they may highlight regions due to increased risk of damage due to vasculature rather than relation-

ship with the outcome measure [44]. A common limitation for both these methods is that they do not scale well for non-binary outcome measures. An example of a subtraction based LSM study by Konczak et al [5] is shown in Figure 2.5.

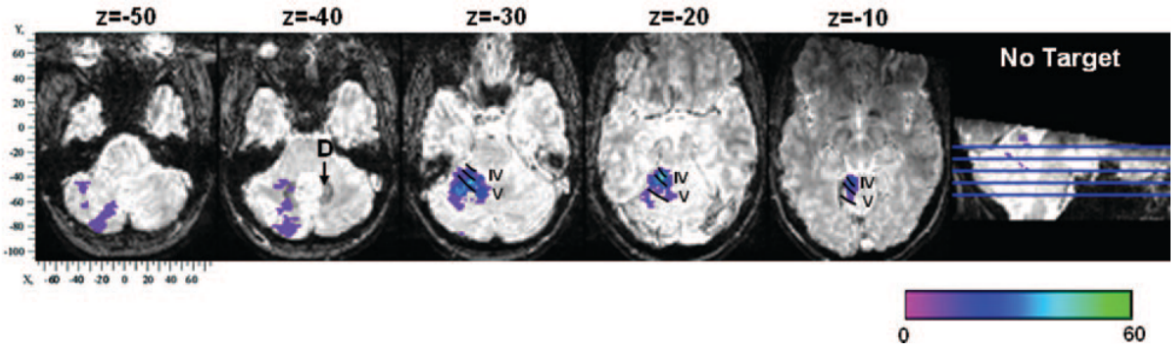


Figure 2.5: Slices of voxel-based lesion subtraction maps. Subtraction maps generated for each voxel by subtracting the percentage of unimpaired subjects with damaged voxel from the percentage of impaired subjects with damaged voxel. Results suggest abnormal kinematic performance strongly associated with damage to the lesions of the cerebral cortex in the paravermal regions of lobulus IV and V. Image taken from the study by Konczak et al [5]

Region of interest

Region-based Lesion-Symptom Mapping (RLSM) have become one of the popular forms of LSM studies. Tradition methods of RLSM require subjects separated by region of damage (unlike overlay which is grouped by outcome measure). A statistical analysis is then performed to assess whether damage to that region is related to some deficit. This statistical analysis may be binary (damage present or not) or continuous (volume of damage to the region). Modern implementations of RLSM use pre-labelled standardised brain atlas which do not require prior grouping of test subjects. Instead this method involves multiple statical analyses to identify statistically significant damage in any of the pre-mapped regions. An advantage of this method is that it is more likely to achieve statistical significance for lower sample sizes than voxel-based methods. A limitation of this method is that the brain regions must be defined prior to analysis and therefore important regions may be overlooked. Related types of analyses are those which use some other metric of brain damage such as the absolute lesion size, although these have lost favour due to knowledge of the importance of lesion location.

Voxel-based

Voxel-based mapping approaches have found widespread use in LSM ever since their novel application in 2003 by Bates et al [6]. Unlike modern implementations of RLSM, this method is capable of statistical analyses on the voxel rather than the region scale enabling unparalleled resolution of analysis. Two of the most common examples of this approach are Voxel-Based Morphometry (VBM) and Voxel-based Lesion-Symptom Mapping (VLSM). The main difference between the two voxel-based techniques is that VBM defines damage to exist on a continuum whereas VLSM considers damage to be binary. Since it is not possible for a human to grade lesion damage in a truly continuous manner, only automated methods of lesion demarcation are suitable for VBM.

Following the demarcation of the affected voxels (either binary or continuous), the next step is to identify whether any statistical significance exists between these voxels and the outcome measure. One of the most common approach with the voxel-based technique is to use a mass univariate approach. This means that each voxel is analysed independently of the others. For VLSM, this involves a t-test performed at each voxel, examining the effect of damage to a particular voxel, across all subjects, with change to a given outcome measure [45]. An example of the results of the first VLSM study conducted is shown in Figure 2.6. This study investigated the brain regions related to speech fluency and language comprehension.

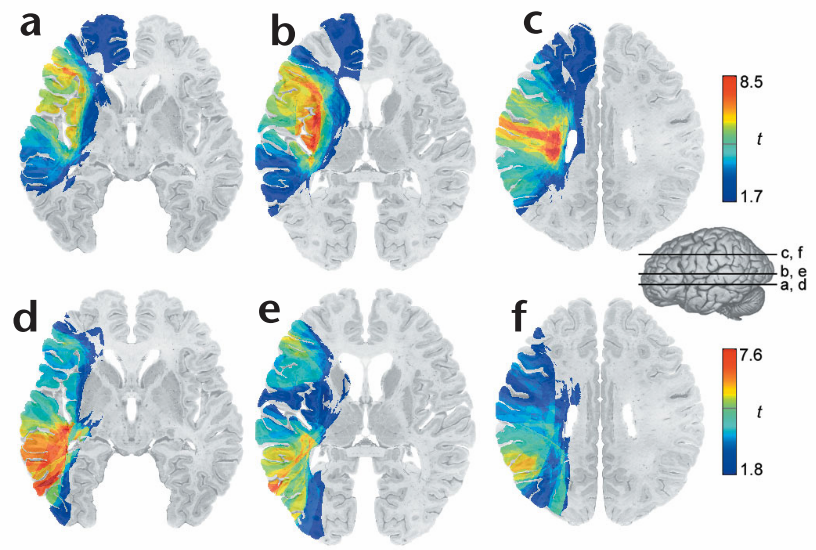


Figure 2.6: Slices of VLSM maps computed for fluency (a-c) and auditory (d-f) comprehension. High t-scores indicating that these voxels have a highly significant effect on the behavioural measure are indicated in red. Taken from [6]

Mass univariate voxel-based methods were once considered state-of-the-art but have lost favour recently due to two main flaws. Firstly, there is the "partial injury problem" [46] whereby if damage to two exclusive brain regions A and B is required to cause a deficit in the motor outcome, then statistical power will be low for both regions and significance may not be found. Secondly, there is the flaw of assuming that each voxel may be assumed to be independent from all other voxels tested. This assumption does not prove valid since brain damage after stroke is systemically distributed across the brain according to the anatomy of the vascular trees [35]. Finally, there is the risk that damage to a brain region may not cause a deficit in motor outcome, but damage tends to coexist with damage to a region of the brain which does cause a deficit. In this case the first region may incorrectly be identified as having a functional role. This is a particularly pressing concern in stroke studies since disruption to the brain vascular system tends to cause patterns of damage which include the same brain regions.

Machine Learning-based Multivariate Models

The aforementioned limitations with voxel-based mass univariate approaches have led to the development of alternative techniques which can study the interaction of multiple brain voxels or regions (as opposed to assuming independence at each voxel). One new technique that has been proposed is the use of supervised machine learning models. These models are trained using features based on the region or voxels of interest and make predictions of a behavioural outcome measure. This approach holds the potential to learn the importance of the interaction between multiple regions in a particular behavioural deficit and as such can be trained to overcome the "partial injury problem" discussed for the mass univariate approach. One limitation of this method is that the direct relationship between lesion location and behavioural outcome may not be clear since this information is learned rather than statistically tested. This is comparison to RLSM or VLSM which can be designed to provide a clear statistical output of the importance of a region/ voxel to a behavioural measure. There are also the limitations which are incurred with any machine learning approach, namely the difficulty identifying the optimal hyperparameters of the model and the risk of overfitting to the data.

2.6 Diagnostic Myography

Myography is the application of devices for the recording of signals analogous to muscle activity in normal and pathological conditions. Common medical applications include the measurement of the onset of fatigue or diagnosis of muscle or nerve dysfunction. An emerging research application has been in the assessment of muscle function changes as a measure analogous to motor dysfunction post-stroke [47].

2.6.1 Electromyography

Electromyography (EMG) refers to the recording of electrical signals generated by motor neurons in skeletal muscles. There are two forms of administering EMG. Intramuscular EMG (iEMG) involves an electrode being inserted into the muscle itself. The method enables a high resolution of measurement but is limited in application by its invasive nature. It is primarily used for the study of deep muscles or muscles which have a small cross-sectional area. An alternative and more commonly used configuration is surface EMG (sEMG). This configuration applies electrodes onto the skin surface and records the muscle activity produced by aggregate muscle fibre firing. This modality is suitable for superficial, large, and easily accessible muscles [48].

2.6.2 Mechanomyography

Mechanomyography (MMG) measures the mechanical low frequency vibration that is produced during muscle contraction [49]. This signal is composed of an initial movement caused by change in muscle shape at muscle contraction followed by lower amplitude vibrations caused oscillations of the muscles fibres at the resonant frequency of the muscle [50]. There as many forms of MMG as there are methods to measure this vibration. Configurations include microphone, accelerometer, and piezoelectric sensor-based. MMG, like EMG, is also best suited for the recording of large superficial muscles.

2.6.3 Application in Post-Stroke Populations

One of the emerging applications of myography is as a diagnostic tool to quantify the motor function in post-stroke populations. A particular application this has been investigated for is in the quantification of motor tasks involving the hand or wrist, a region which

is notoriously difficult to measure using more conventionally applied inertial sensing. Although neither surface EMG nor MMG is truly capable for measuring the individual small muscle groups responsible for fine motor tasks, there is evidence that these movements may still be detected from the aggregate muscle activity, and that a measure of motor dysfunction may be determined [15].

Myography offers the potential to provide new insight into the motor dysfunction of these regions than the current purely visual or kinematic procedures can offer. EMG holds an advantage over MMG for this application in terms of being more widely validated for diagnostic applications. Conversely MMG possesses the advantages of re-usability, possessing a robust signal which is more independent to changes of skin impedance, and ease of application. Both modalities suffer from motion artefacts and the presence of crosstalk, whereby unwanted signals are picked up from adjacent or deep muscles. Studies involving these sensors must be designed with configurations and signal processing methods to minimise these errors.

2.7 Motion Tracking

Motion-based sensors have continued to decrease both in size and cost over the past few decades. This factor combined with unheralded advances in computer power and data analysis algorithms have meant that motion capture sensors are being used more widely than ever before. This includes their newfound application in a healthcare setting, including pervasive monitoring and as a substitute or complement to the standard clinical rating scales. Sensors used in motion tracking may be broadly categorised as wearable motion capture sensors, which instrument the body of interest, and visual motion capture sensors, which use visual means to record motion remotely from the body of interest.

2.7.1 Wearable Motion Capture

The simplest form of wearable sensor is composed of a single accelerometer or gyroscope. Accelerometers operate by measuring the acceleration present on a body in the local axis frame. This reading will include the presence of the gravity vector which may be utilised or filtered out of the subsequent data. Gyroscopes measure the angular velocity present on a local axis frame. Both sensor types may be single-axis or tri-axial, measuring values in or around each of the three principal axes.

More complex wearable systems may include an inertial measurement unit (IMU). This system is composed of both a triaxial gyroscope and a triaxial accelerometer. In addition, the IMU may also include a triaxial magnetometer. The function of the magnetometer is to measure a magnetic field vector (typically the earth's magnetic field). This vector coupled with the gravity vector measured by the accelerometer provide two global vectors from which a global orientation may be calculated. A diagram of each of the sensors of the IMU is shown in Figure 2.7.

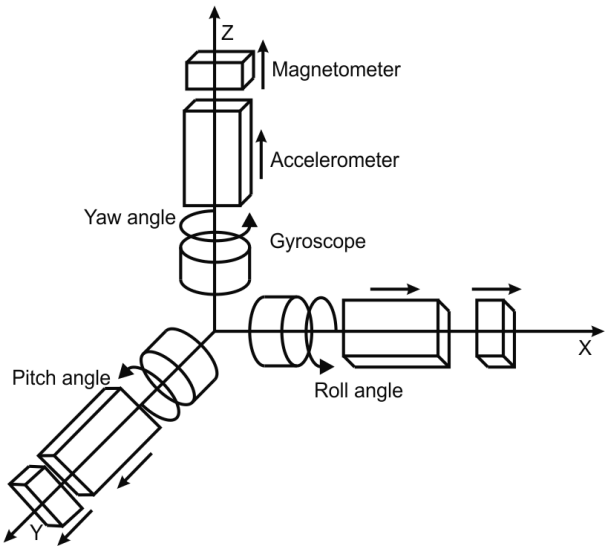


Figure 2.7: Components of an IMU. The linear components of accelerometer and magnetometer measure local changes as well as the presence of global vectors. The gyroscope measures local rotation around each of the axes. Taken from [7]

2.7.2 Visual Motion Capture

Visual motion capture systems utilise two or more cameras to triangulate the 3D position of an object. The traditional and gold-standard method of detecting body locations is using markers attached to the subject. These markers may then be detected and accurately tracked by cameras while in the field of vision. Several recent systems have achieved methods of dynamically tracking surface features of the subject without the requirement for markers. The method has raised the potential for visual motion capture systems to be used in a clinical environment by reducing the risk of cross-contamination and lowering the setup time required for capture.

The Vicon system (Vicon Motion Systems Ltd., Oxford, UK) is one of the most

sophisticated commercially available visual motion capture systems. Like other high-end systems, the Vicon requires numerous cameras spaced around a room to be able to fully capture an individual wearing specially designed markers. The Vicon system has been applied in a variety of research applications due to its high level of accuracy. However, the high cost and long setup times mean that it is not always feasible to apply such systems in a clinical environment.

One of the most commonly available visual motion capture systems is the Kinect system (Microsoft Corporation, Redmond, WA). The Kinect system uses a VGA video camera combined with a depth sensor to build up a human skeleton model in 3D space. The ability to function without the requirement for markers, ease of setup, and ease of extracting sophisticated orientation data have enabled this system to be commonly implemented in research. Although this system may not achieve the same level of accuracy offered by more sophisticated visual motion capture systems, this is compensated by the low cost and convenience of use in applications where very high accuracy is not paramount.

2.8 Defining Orientation

A key component of motion tracking is the interpretation of sensor data as an orientation in 3D space. The quantification of orientation in a human body enables more comprehensive information about human motion than may be discerned from raw sensor data alone. The includes information about posture, gait, and limb joint angles. This information also has clinical relevance in a healthcare environment. Common clinical metrics which may be derived from orientation information include a patient's range of motion or joint synergies for instance. Two of the most common ways of defining orientation in a 3D space are by Euler angles and quaternions respectively.

2.8.1 Euler Angles

Euler angles define the orientation in a 3D space by the composition of three successive rotations around an axis frame. These rotations may be extrinsic, rotation around a fixed axis frame, or intrinsic, in which the reference axis frame rotates along with each successive rotation. Euler angles have the useful property of being simple to visualise but suffer the limitation of Gimbal Lock. This occurs when the orientation cannot be

uniquely defined by a set of successive rotations and results in an unstable output. An example of an extrinsic XYZ (defined as rotation in the order of x-axis, y-axis, and then z-axis) is shown in Figure 2.8.

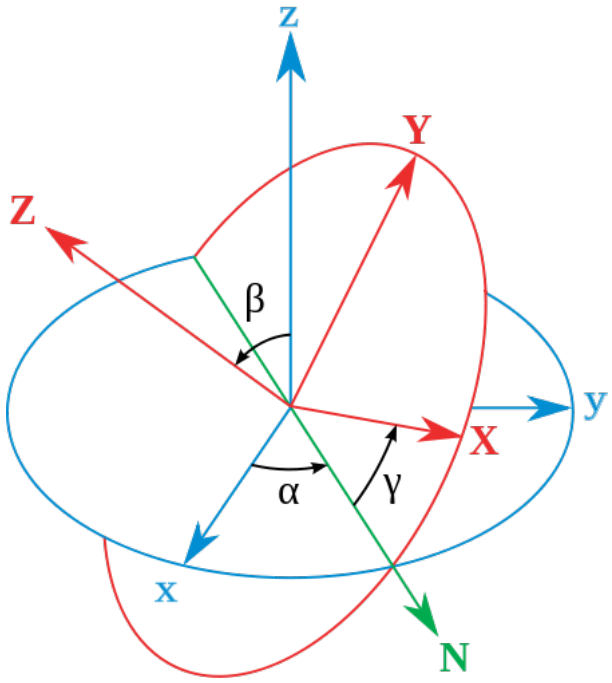


Figure 2.8: Extrinsic Euler angles defined in the XYZ order

2.8.2 Quaternions

Quaternions are an alternative method of defining orientation. Unlike Euler angles, quaternions require four dimensions to define a body in the three-dimensional space. Three of these components are purely imaginary and one component is real. Quaternions possess the advantage over Euler Angles of not suffering from Gimbal Lock since every possible orientation may be representing uniquely by a quaternion and its negative representation. Another advantage the quaternion form has is the ease of multiplication between quaternions or vectors to represent successive rotations or vector rotation. This will be discussed in more depth later in this section. These useful properties have led to quaternions becoming the preferred means of quantifying orientation in a range of applications including animation, video games, and research.

Quaternion Representation

Quaternions are a number system which extends the definition of complex numbers and as such the rules which apply for this system also apply for quaternions. The four-dimensions of a quaternion consist of one real dimension and three imaginary dimensions. The general form is expressed as:

$$q = a + bi + cj + dk$$
$$where : i^2 + j^2 + k^2 = ijk = -1 \tag{2.1}$$

The quaternion may be represented in its purely imaginary (3-element vector) form or its purely real form (single element). These representations will be denoted in this thesis as follows:

$$q_{Im} = Im(q) = Im(a + bi + cj + dk) = [b, c, d] \tag{2.2}$$
$$q_{Re} = Re(q) = Re(a + bi + cj + dk) = a \tag{2.3}$$

The quaternion norm is calculated in the way that would be expected for a 4-element vector. This is an important concept since it can be used to find the normalised quaternion which makes subsequent calculations much simpler. The method for calculating the norm and finding the normalised quaternion respectively are given below:

$$|q| = \sqrt{a^2 + b^2 + c^2 + d^2} \tag{2.4}$$
$$q_{norm} = \frac{q}{|q|} \tag{2.5}$$

The inverse quaternion is an important concept since it represents the inverse rotation. This concept is required for taring quaternions as well as in vector and quaternion multiplication. The inverse is equal to the conjugate for normalised quaternion since the quaternion norm is equal to 1 in this case. The conjugate and inverse quaternion equations are displayed below:

$$q^* = a - bi - cj - dk \tag{2.6}$$

$$q^{-1} = \frac{q^*}{|q|^2} = q^*(\text{for normalised quaternions}) \tag{2.7}$$

Quaternion Multiplication

Quaternion multiplication is non-commutative, and the product produces a quaternion which represent two successive rotations in the same 3D space. In this thesis, multiplication between two quaternions will be represented by the symbol $\otimes$. The equation for multiplication between two quaternions is given below:

$$q_1 \otimes q_2 = (a_1 + b_1 i + c_1 j + d_1 k) \otimes (a_2 + b_2 i + c_2 j + d_2 k)$$

$$\begin{aligned} q_1 \otimes q_2 = {} & (a_1 a_2 - b_1 b_2 - c_1 c_2 - d_1 d_2) \quad + \\ & (a_1 b_2 + b_1 a_2 + c_1 d_2 - d_1 c_2) i \quad + \\ & (a_1 c_2 - b_1 d_2 + c_1 a_2 + d_1 b_2) j \quad + \\ & (a_1 d_2 + b_1 c_2 - c_1 b_2 + d_1 a_2) k \end{aligned} \tag{2.8}$$

Quaternion multiplication may also be used to apply a rotation to a three-dimensional vector. For this method, the vector should be formatted as a pure quaternion with zero real scalar component and the vector as the imaginary components. This can then by multiplied between the quaternion and the inverse form of the same quaternion to apply the rotation. The method for rotating a vector (v) by a quaternion (q) is shown below:

$$\begin{gathered} v = [v_1, v_2, v_3] \\ q_v = quat(v) = [0 + v_1 i + v_2 j + v_3 k] \\ q_{rot} = q \otimes q_v \otimes q^{-1} \\ v_{rot} = Im(q_{rot}) \end{gathered} \tag{2.9}$$

Quaternion Swing-Twist Decomposition

Quaternion swing-twist decomposition is a method of decomposing a quaternion into two concatenated quaternions: swing and twist. Given a twist axis (v_T), this operation can be used to calculate the quaternion which represents the portion of the rotation that only

defines the twist around this axis (q_{twist}). The remaining quaternion represents the swing portion (q_{swing}) from the defined twist axis. There are several methods of deriving this output, but the operation selected for this project is as follows. An intermediate pure quaternion (q_{v_T}) represents the rotation of the twist axis by the original quaternion. The variable (p) represents the projection of the twist axis.

$$q = a + bi + cj + dk$$

$$v_T = [v_1, v_2, v_3]$$

$$q_{v_T} = q \otimes quat(v_T) \otimes q^{-1}$$

$$p_1 = cross(v_T, Im(q_{v_T}))$$

$$p_2 = dot(v_T, Im(q_{v_T}))$$

$$q_{swing} = [1 + p_2, p_1[0], p_1[1], p_1[2]] \tag{2.10}$$

$$q_{twist} = q_{swing}^{-1} \otimes q \tag{2.11}$$

2.8.3 Quaternion Estimation Algorithms

A number of algorithms have been developed to estimate an orientation based upon the calculation of some combination of different sensor metrics. The traditional method of estimating orientation from potentially noisy sensor data has been the use of the Kalman filter. A more recent method of orientation estimation that uses the gradient descent optimisation procedure will also be discussed.

Kalman Filter

A Kalman filter [51] is an algorithm which uses the combination of statistically noisy measurements to achieve a prediction of an unknown variable which is more accurate than possible using any single measurement alone. This is achieved through the use of an estimated joint probability distribution over the variables for each timeframe. The Kalman filter is particularly well suited for orientation estimation applications due to the simplicity of its mathematical derivations and its recursive nature [52]. Non-linear

implementations of the Kalman filter have been used for orientation estimation, most commonly the extended Kalman Filter [53] and Unscented Kalman Filter [54].

Madgwick Gradient Descent

The Madgwick Gradient Descent algorithm is a novel orientation algorithm which uses a low-computational load and can be adjusted to perform well under noisy sensor data. It utilises the gyroscope measurement to filter out high-frequency errors whilst the accelerometer, and optionally the magnetometer, are used to deal with the integral drift. The full derivation of this algorithm may be found in this paper by Madgwick et al [55] but a brief overview is discussed below.

Equations for the differential and non-differential quaternions involving angular velocity may be estimated using the equations below. The differential quaternion equation may be satisfied by an initial guess of the quaternion, and an updated value of angular velocity. The second equation may be satisfied using the result of the prior equation in addition to the time period.

$$\dot{q_{w,t}} = \frac{1}{2} q_{w,t-1} \otimes w_t \tag{2.12}$$

$$q_{w,t} = q_{w,t-1} + \dot{q_{w,t}} \Delta t \tag{2.13}$$

The quaternions (q) in both the above equations represent estimations of the earth's reference frame relative to the sensor reference frame. The gyroscope reading (w) is defined in the equations as a pure quaternion to enable quaternion multiplication. The quaternions in the equation below also contain the subscript (w) to represent that they have been derived from the measure of angular velocity.

Gradient descent optimisation is used for the estimation of a second quaternion from the accelerometer and magnetometer readings. The objective function for the optimisation is derived by minimising the difference between a reference vector (d) in the earth frame rotated into the sensor frame and the same reference vector in the sensor frame (s). The subscript (∇) present in the following equations denotes that the quaternions have been derived by gradient descent.

$$\min_{q \in \Re^4} f(q_\nabla, d, s)$$

$$f(q_\nabla, d, s) = q_\nabla \otimes d \otimes q_\nabla^{-1} - s \tag{2.14}$$

This objective function and the respective Jacobian of the optimisation problem simplify considerable when you consider that the acceleration (and magnetometer) reference vectors in the earth reference frame are not present in each of the principal axes. For instance, the gravity vector can be considered to only be active in the z-axis while the earth's magnetic field is only present in two axes.

Finally, there is a fusion step whereby the information provided by the quaternion derived from the gyroscope and the quaternion derived from accelerometer and magnetometer are combined. Both quaternions are important since: (1) the quaternion derived from the acceleration and magnetometer readings is based on an initial "guess" measurement, and (2) the quaternion based on the gyroscope will accumulate errors due to sensor noise. The ideal fusion algorithm should filter out the high-frequency errors in (1) while compensating for the integral drift subject to (2). The method for optimising the weights of how these quaternions should be combined is beyond the scope of this thesis but the general form of the equation is as follows:

$$q_t = \gamma_t q_{\nabla,t} + (1 - \gamma_t) q_{w,t}, \quad where \quad 0 \leq \gamma_t \leq 1 \tag{2.15}$$

The Madgwick Gradient Descent algorithm has been proposed as a less computationally intensive version of the Kalman filter which makes it more suitable for onboard computation or for use in large scale analyses (as used in the present study). The performance of this algorithm has also been benchmarked at comparable values to the Kalman filter with $< 0.8^\circ$ static RMS error and $< 1.7^\circ$ dynamic RMS error.

2.9 Machine Learning in Healthcare

The exponential growth in data available combined with advances in computer power and algorithm development has led to unprecedented interest in the application of machine learning. This is particularly true in the healthcare fields due to several driving forces. This includes the potential of numerous different healthcare applications which are well suited for automation, such as the use of machine vision in assessment of patient scans or analysis of sensor data for monitoring patients. Another factor is the increasing demands placed on healthcare in many countries due to ageing populations. This has increased the demand for automation as a means of reducing the workload placed on clinicians. This

section will outline the machine learning practices required for any high-quality machine learning study in healthcare.

2.9.1 Feature Selection

Feature selection is an important step in the design of any machine learning study which possesses a feature set which exceeds the number of data instances. The objective of feature selection is to remove features which either do not improve the predictive model or highly correlate with other existing features. One advantage of feature selection is that it reduces model training time since a reduced subset of features are required for training. Another advantage is a reduction in the risk of over-fitting due to the reduced variance and increased generalisability properties of a simpler model. The two main forms of feature selection are filter methods and wrapper methods.

Filter methods of feature selection compute a ranking list of features based on their perceived usefulness. Once a ranking has been determined, the top x features may be selected for further analysis. A common filter method is the use of a correlation measure such as the Pearson's correlation coefficient. The method ranks features based on high a correlation value is achieved between each feature and the labels. Another common filter method is the Relief algorithm. This method uses a distance metric to find the nearest same and different class instances for a series of random instances. The distance of these instances is used to compute a weighting. A high weighting is assigned for close same class instances and far different class instances to reward features which achieve high inter and low-intra class separability.

Wrapper methods utilise a form of classification model in the feature selection process. This is typically permutation-based whereby a set of features is constantly evaluated and modified to find the feature set which provides the optimal classification performance. A common wrapper method is the greedy forward search whereby classification performance is tested with a single feature, and then additional features are only included to the set if they result in a notable improvement in performance accuracy. Another wrapper method is the exhaustive search. This method calculates the classification performance for every single possible combination of features in the set and then selects the combination which results in the best performance. This method is more likely to find the optimum combination of features than the greedy forward search but leads to an impractical number of iterations for high dimensionality features spaces.

2.9.2 Dimensionality Reduction

Dimensionality reduction, like feature selection, may also be used to reduce the number of features in the dataset. Dimensionality reduction offers the same advantages of faster training time and reduced risk of over-fitting. Unlike feature selection, features are transformed into a lower dimension and do not retain their original values. Two of the most frequently implemented examples of dimensionality reduction are Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA).

PCA is used to transform the original feature space to a new space composed of perpendicular vectors which explain the variance of the data. An advantage of PCA is that the vectors returned lie perpendicular to each other and therefore do not correlate with each other. Another advantage of PCA is that a subset of the transform vectors may be selected while still explaining the vast majority of the variance of the original dataset. In this way, PCA may be used as a simple way to reduce the number of features while still preserving the data variance. An example of PCA used to decompose the eigenvectors of a multivariate Gaussian distribution in 2-dimensional space is shown in Figure 2.9. The two main methods of performing PCA are by singular value decomposition of a design matrix and by eigenvalue decomposition of the covariance matrix.

LDA, like PCA, also attempts to lower the dimensionality of the feature space by projecting the data to a lower dimension. Unlike PCA, LDA also uses information about the classes assigned to samples to maximise class separability. Standard implementations of LDA assume that each feature follows a Gaussian distribution. Statistical measures are then computed for the features to find which features maximise class separability and minimise intra class variance. Once this has been computed, a function may be developed to project the feature data into this lower feature-space.

2.9.3 Regression

Regression models have been developed to map real input values (independent variable) to predict continuous target values (dependent variable). One of the simplest forms of regression, the linear regression model, assumes that the target values are a linear composition of input values multiplied by a series of coefficients. The most common form of determining the optimal coefficients for linear regression is by ordinary least squares. This method seeks to find the best-fitting line/ plane by minimising the sum of the squared residuals.

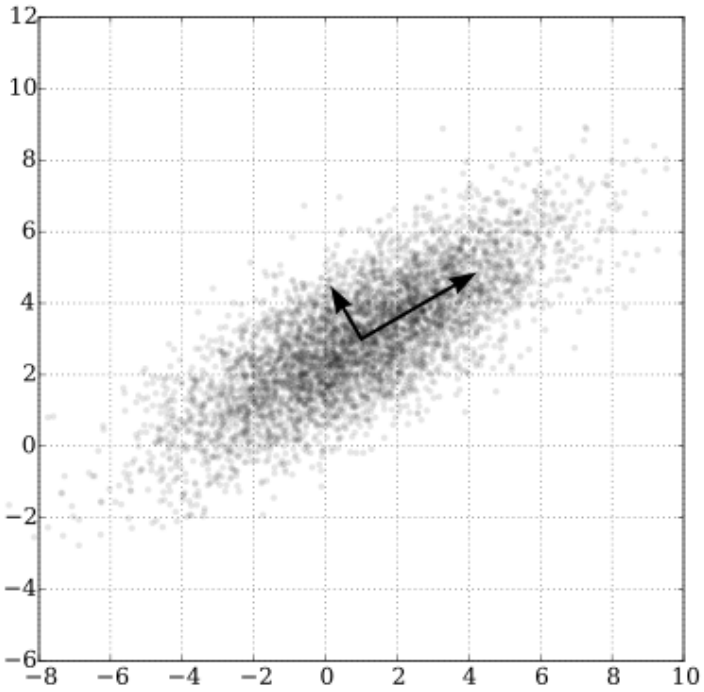


Figure 2.9: PCA of a multivariate Gaussian distribution in 2-dimensions. The 2 eigenvectors are illustrated by arrows. The first eigenvector explains the majority of the variance of the distribution

Regularisation may also be applied to the regression problem to reduce the complexity of the model and therefore reduce the risk of over-fitting. Common examples of regularisation include Least absolute shrinkage and selection operator (Lasso) and Ridge regularisation which minimise the absolute sum and the squared absolution sum of the coefficients respectively. Since Lasso regularisation uses the absolute sum of coefficients, it will also drive low coefficients down to zero and therefore also acts as a feature selection step.

2.9.4 Classification

Classification models use a function to map real input values to predict discrete target values. In this subsection, the function of a few of the most common classification models will be discussed.

Decision Trees

Decision trees at their simplest form one of the most intuitive classification models. Labels (leaves of the tree) are selected based on whether certain feature threshold values are met at each of the branches of the decision tree. Many different methods exist to attempt to

train the optimal threshold values as well as the overall complexity of the tree to achieve maximal class separability. A pruning step is also frequently implemented to remove unnecessary branches of the decision tree and in turn combat over-fitting.

K-Nearest Neighbours

The K-nearest neighbours (KNN) algorithm selects the target value based on the K nearest training instances to a new test instance in n-dimensional space. For this model, K refers to the number of neighbours to utilise, and n refers to the total number of features. The K nearest neighbours to the test instance are calculated using a distance measure, such as the Euclidean distance. A vote is then performed based on the classes of the K nearest neighbours with a majority vote representing the class prediction. A representation of KNN with class assigned based on the 5-nearest neighbours is shown in Figure 2.10.

Support Vector Machines

The support vector machine (SVM) is an algorithm which seeks to find a hyperplane in n-1 space (where n is the number of features) which maximally separates the support vectors representing each class. In this case, the support vectors are instances in n space which lie close to the hyperplane and therefore will influence the position and orientation of the hyperplane. Maximal margins refer to the maximum amount of separation achieved between the hyperplane and the support vectors. An example of the application of SVM to separate a 2-class problem in 2-dimensional feature space is shown in Figure 2.11

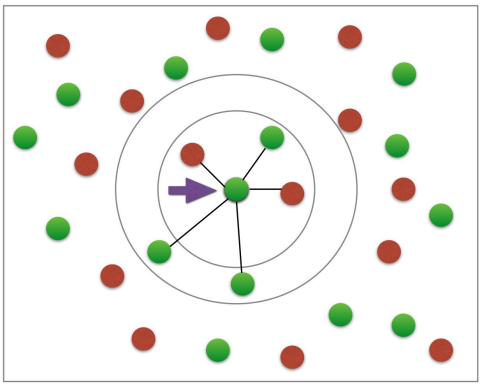


Figure 2.10: Representation of KNN applied in 2-dimensional space with K set to the 5 nearest neighbours

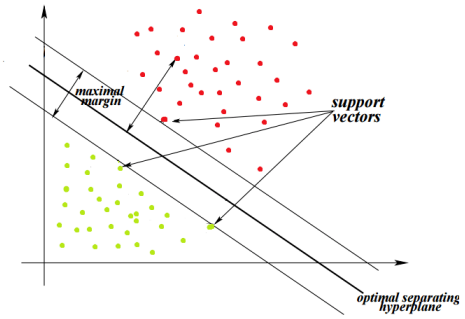


Figure 2.11: Representation of SVM hyperplane in 2-dimensional space showing the support vectors and maximal margin

Artificial Neural Networks

An in-depth description of artificial neural network (ANN) is beyond the scope of this thesis. However, a fundamental description is outlined here to provide context to similar studies which have implemented these models. At their simplest ANNs consist of an input layer, hidden layer, and output layer. Features are provided as the input layer, and coefficients provided in the hidden layer dictate how much influence these have on selecting the output (class) to select. More complexity is provided by including more hidden layers. The optimisation of the coefficients provided in the hidden layer depends on the type of neural net selected. ANNs can achieve unparalleled performance for large datasets but may be outperformed by other classification methods for small datasets.

2.9.5 Ensemble Decision Tree Classification

Ensemble learning algorithms determine an output based on the result of numerous learner models. This aggregate method may use predictions from different classification models, feature sets, or samples for instance. This method has been shown to have the potential to achieve a more accurate classification than possible by a single learner model alone [56].

Bagging

Bagging is an ensemble method which uses bootstrapping (sampling from a set with replacement) to enable a fully random selection of instances provided for each classifier. For each iteration, a selection of bootstrapped samples is selected and fed into a classifier. This is then repeated for n-1 iterations and the collection of predictions are aggregated to select the output with the most votes. An image of the bagging method as applied to a series of decision trees is shown in Figure 2.12.

Random Forest

The random forests ensemble method, like bagging, also implements bootstrapping to provide random instances to a series of classifiers. Unlike bagging, each decision tree is trained on a unique and random subset of the total feature set. This results in a series of decision trees which will develop branches in a unique way depending on the available features. This method is typically more robust to over-fitting than bagging methods. The general idea of the structure of the random forest method is shown in Figure 2.12.

However, the decision trees would be non-identical either in terms of the branches formed or the thresholds at each branch due to unique feature subsets.

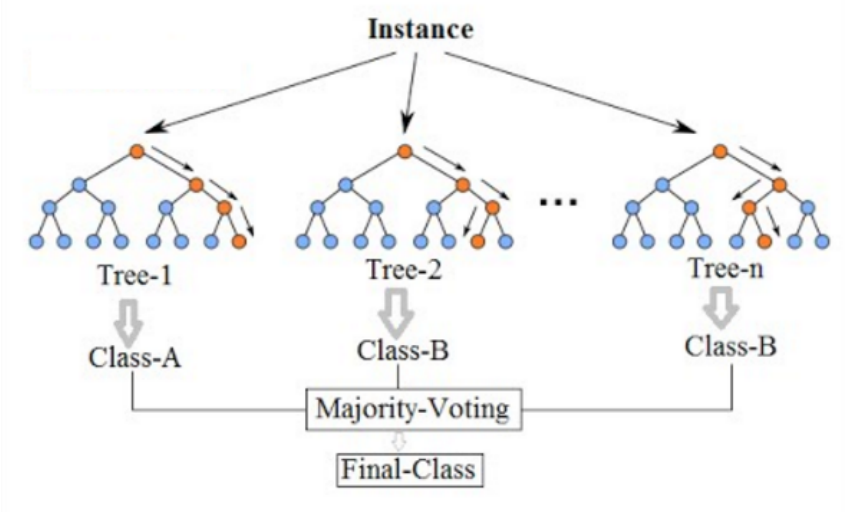


Figure 2.12: Diagram of the structure and prediction of bagging/ random forest type ensemble classifiers. Taken from [8]

Boosting

The boosting ensemble generic algorithm, like the bagging method, uses bootstrapping to sample instances from a set and train a series of weak learners. The primary distinction between the two methods is how the weak learners are trained. For boosting methods, weak learners are trained in series, rather than parallel, which means that subsequent weak learners in the chain are dependent on the previous. Another notable property of the boosting algorithm is that each of the weak learners is assigned their own weighting based on how well they classify unknown instances in the training stage. This weighting is subsequently used at the testing stage whereby higher weight is assigned to the weak learners which performed better in the training stages. An example of the boosting algorithm, whereby the weakness in a learner can be seen to update the parameters of each subsequent learner, is shown in Figure 2.13.

AdaBoost

The AdaBoost (Adaptive Boosting) algorithm is a form of boosting algorithm which modifies the sampling of instances by assigning weights to each instance, which affects its likelihood of being sampled. A subset of these weights is modified on every iteration in

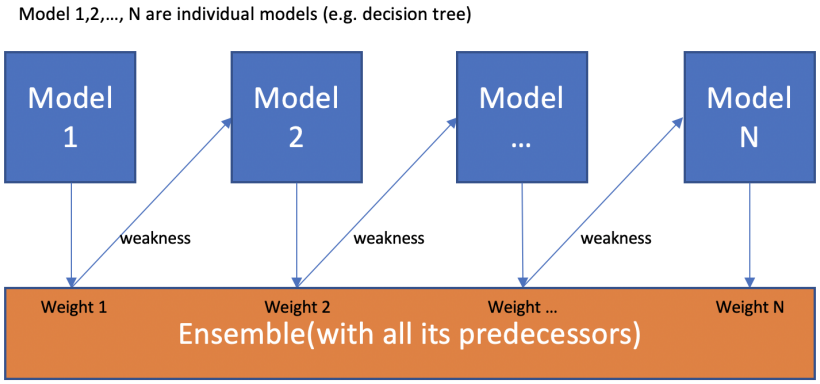


Figure 2.13: Boosting algorithm structure whereby subsequent models are trained based upon the weaknesses of their predecessor. Taken from [9]

which a new decision tree is trained, based on how well instances were classified during the training step. Instances which are poorly classified are assigned a large weight, meaning future decision tree models which implement them in their training data and improve the ability of the ensemble model to classify these instances.

Gradient Boosting

Gradient Boosting algorithms are another form of boosting algorithm. Unlike the AdaBoost algorithm, performance is modified iteratively using a loss function rather than by applying weights to the data instances. This loss function is a measure of the misclassification such as squared error for regression or logarithmic loss for classification. A gradient-descent like method is used to move towards the minimum amount of loss for each new decision tree generated by modifying certain classification parameters each iteration.

Light Gradient Boosting

Light gradient boosting (LightGBM) is a form of gradient boosting which has gained popularity since its release due to its superior performance and speed compared to other gradient boosting algorithms. The property which has led to this improvement is the unique way the LightGBM classifier grows its trees. Whilst typical decision trees grow by levels up until a certain constraint, the LightGBM classifier grows each decision tree leaf-wise. This means that it can achieve sufficient model complexity to classify certain labels while minimising model complexity and reducing the risk of over-fitting when necessary. A diagram of how this algorithm grows the model leaf-wise (as compared to standard level-wide growth) is shown in Figure 2.14.

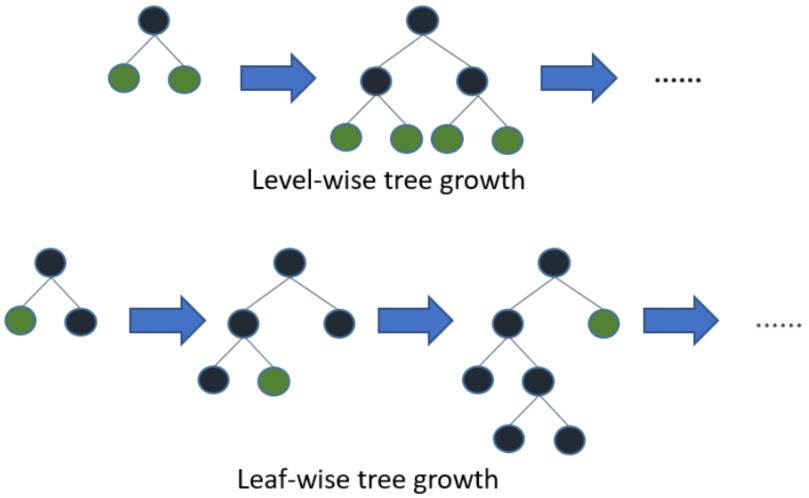


Figure 2.14: An example of leaf-wise decision tree growth (as opposed to level-wise in which all branches are grown equally across the level) as implement in the LightGBM algorithm. Taken from [10]

Performance Measures

An important factor in any classification analysis is how the result is presented. A telling example of this is if for a two-class problem the vast majority belongs to one class, then a classifier may achieve high accuracy by continually predicting one class only. In this case, the classifier may appear to be performing well when in reality the classification is trivial. To prevent over-estimation of classifier performance or to assess different parameters of the classifier, there are several other commonly used performance measures other than just accuracy. Equations for common performance metrics are given below for binary classification problems (although they may be extended for multi-class problems). An example of a binary confusion matrix is shown in Figure 2.15.

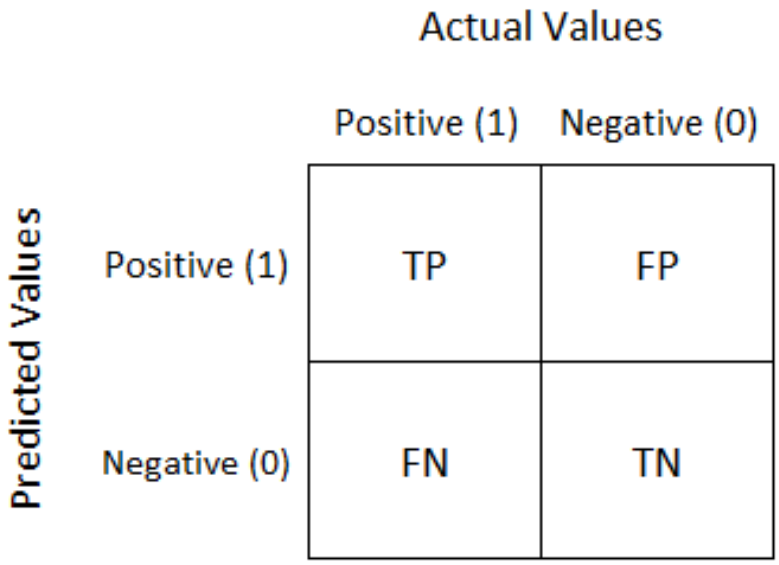


Figure 2.15: Confusion matrix for a binary classification problem

The accuracy of a classifier is the most straight-forward measure of classifier performance. It gives a good indication of classifier performance but may give misleading results if the data is biased or if false positives are particularly undesirable in the analysis:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{2.16}$$

Recall is a measure which detects how often a positive result is detected in the data. This is particularly important in analyses which detection of a measure is critical:

$$Recall = \frac{TP}{TP + FN} \tag{2.17}$$

Precision measures how often the classifier erroneously gives a positive result. This may be more important in analyses where false positives are undesirable:

$$Precision = \frac{TP}{TP + FP} \tag{2.18}$$

Finally, the F1-score is a common metric used in classification analyses which integrates both the recall and the precision. This results in a score which is more sensitive to false

positives and false negatives than the accuracy score.

$$F1\text{-}score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{2.19}$$

CHAPTER

THREE

Wearable Fusion System for Automated Rating of Upper-Arm Motor Function Post-Stroke

3.1 Chapter Introduction

Traditional quantification of post-stroke motor function is performed by clinicians using one of the standard clinical rating scales. These rating scales are well-validated and widely used but suffer from drawbacks including low resolution and high subjectivity. The work conducted in this chapter aimed to develop a wearable sensor system which combined the validity provided by the standard clinical rating scores, with a number of novel sensor metrics which are not limited by the low accuracy and resolution of manual assessment measures.

The system was designed with a combination of kinematic sensors in the form of IMUs, and myographic sensors in the form of MMG sensors. The kinematic data enabled the calculation of advanced orientation features and the development of an avatar for online/ offline assessment purposes. The inclusion of myographic sensors enabled measurement of hand and wrist function which is otherwise difficult to quantify using the inertial- and camera-based sensors conventionally used in wearable systems.

A prediction of clinical score was achieved by feeding features derived from the kinematic and myographic sensor data into a specially developed classification pipeline. A selection of fine-grained clinical features was also calculated to complement the predicted score and provide more in-depth information about the sub-components of motor dysfunction.

The quality of the system was determined both by the performance of the classifier

(as compared to prior studies) and the usefulness of the clinical features (as determined by correlation with the clinical rating score). A system which can be evidenced to provide good measures of both these metrics has the potential for use in both clinical applications and research studies.

3.2 Chapter Outcomes

The novel fusion system developed in this study was used to produce two major outcome measures. Further clarification of these metrics is stated here to provide justification for these outcomes as well as the application in the second study included in this thesis.

The first metric produced is a prediction of the FMA-UE clinical rating. This metric is produced using a boosted machine learning model which makes predictions of clinical score using new sensor data, based upon historic sensor data as well as previously assigned clinical ratings. This metric type forms the predominant output of most prior studies of wearable systems post-stroke.

The second metric produced by the system is a series of clinically relevant fine-grained features. Features were selected from those which had already been generated for the above machine learning model based upon a number of useful criteria. The justification for including these features in isolation (as opposed to just feeding them into the above machine learning model) is that the resolution of these features is limited only be the resolution of the sensors themselves as such far exceeds that of the FMA-UE. If these features can be proven to be intrinsically useful measures of motor dysfunction then they may be used in isolation of or to complement the FMA-UE as a more "pure" measure of motor function. Due to the high-resolution nature of these clinical features (as compared to the FMA-UE), they were introduced into the subsequent lesion symptom mapping study covered in the next chapter.

3.3 Chapter Structure

This chapter is structured as follows:

- **Literature Review**- Review of the instrumentation and major processing stages of prior systems developed for automated scoring of motor function. Prior methods

for determining motor outcome metrics (generally be machine learning models) will also be discussed.

- **Instrumentation**- A description of the in-house instrumentation included in the wearable system proposed for this study. Described in terms of sensor function, configuration, and sampling
- **Software**- The novel software developed for this project comprised of GUIs for the data collection, data processing, and classification stages
- The fundamental stages of the study in terms of:
  - **Data Collection**
  - **Data Processing**
  - **Data Visualisation**
  - **Feature Extraction**
  - **Feature Exploration**
  - **Classification Pipeline**
  - **Clinical Feature Set**
- **Results**- Described in terms of the classification performance of the developed machine learning model, and the calculated usefulness of the derived clinical feature set
- **Discussion**- The overall results of the project discussed in terms of meeting the objectives of the study and how they compare to the prior systems described in the literature review.

3.4 Literature Review

This literature review is composed of prior studies which have incorporated sensors into a system either for the prediction or correlation with a score of motor function in post-stroke subjects. These systems will collectively be described as "automated" systems in this thesis. The focus of this review will be on the upper-extremity motor function, although studies which examined lower-extremity motor function will also be discussed in less depth.

3.4.1 Instrumentation

The instrumentation used in prior automated systems of motor function post-stroke may be broadly categorised into four types. Firstly, **wearable motion capture systems** are composed of sensors placed on the body segments of the subject being tested and detect segment activity directly. Secondly, **visual motion capture systems** which use visual sensors to detect body segments and calculate body metrics such as acceleration and joint angle indirectly. Thirdly, **robotic motion capture systems** which require the subject to perform movements inside a robotic device. Finally, an emerging modality which has only recently been trialled in systems of motor function post-stroke are **myographic capture systems**, which use sensors to record measurements analogous to muscle activity.

The instrumentation of most relevance to the present study are those related to wearable/ visual motion capture and myographic capture systems. These devices will be discussed in greater depth within this literature review, including in terms of motor features which have been calculated from the sensor data in Subsection 3.4.4 Feature Calculation.

Wearable Motion Capture

The use of accelerometers forms the simplest wearable setup adopted for automated motor function test post-stroke. Accelerometers placed on limb segments offers a low-cost automated assessment which can provide a usable signal with minimal processing. The measurement of the gravity vector may be utilised to predict the arm orientation or may be removed entirely by high-pass filtering of the signal. Disadvantages of this setup include the lack of any angular velocity or orientation information which would allow more complex interpretation of task performance. The low fidelity of information

provided by this modality compared to alternative sensors has meant that it has become less frequently utilised as a sole measure in clinical systems of motor function. However, its low power consumption has meant that it is still being trialled in systems of pervasive tracking of motor recovery [57, 58].

One of the earliest systems developed for measuring post-stroke motor function was the 2005 study by Knorr et al [59]. The system proposed by Knorr et was composed of dual-axis accelerometers placed on the hand, forearm, and upper arm. The study was then extended the following year [60] to include single-axis accelerometers placed on the fingers of the hand, presumably to better capture hand/grasp tasks. The same system was utilised again in 2010/ 2011 on a larger population and with a more complex analysis by Patel et al [16] and by Del Din et al [17]. An image of this system is shown in Figure 3.1. The most recent example of an accelerometer only system was the 2014 study by Wang et al [61]. The study by Wang et al developed a system comprised of two tri-axial accelerometer nodes placed on the forearm and upper arm. The application of tri-axial accelerometers provides information in each local axis (unlike the aforementioned studies) but the absence of any hand or wrist sensors means there is no information provided concerning this region.

The introduction of IMU systems in wearable sensing systems has heralded a big advance in the fidelity of information collected. This is due to the inclusion of a tri-axial gyroscope (and possibly a magnetometer) sensors in addition to the accelerometer. Recent developments in computational power and orientation algorithms (Kalman filter, Madgwick Gradient Descent) have also made it possible to fuse the aforementioned modalities to produce an approximate orientation of the IMU system relative to a fixed axis frame. This may be used in isolation to calculate limb orientation or in relation to other limb segments to compute joint angles.

A common alternative to wearable sensing for capturing kinematic data is by visual motion capture, of which the Kinect system has been used most frequently for assessment of post-stroke motor dysfunction. Compared to the Kinect system, systems implementing a number of IMUs possess several significant disadvantages (limitations of the Kinect compared to IMUs are detailed in Subsection 3.4.1: Visual Motion Capture). Two such disadvantages are their relative expense and battery life. Each IMU system may cost as much as a few hundred pounds and several may be required for fully instrumenting a section of the body. This is compared to the Kinect sensor which originally retailed for approximately £100. Secondly, each IMU involves collection of data from several sensing

modalities which requires a significant power draw. In addition, each IMU typically requires its own power source (battery) meaning that they need to be recharged often. This is in contrast to the Kinect sensor which draws its power from a laptop, which in turn may be connected directly to a power supply. A final major limitation of the IMU system is noise and drift incurred during orientation calculations.

Orientation drift has historically been a common problem when computing orientation from wearable sensors due to an accumulation of small errors associated with the gyroscope. Advances in orientation algorithms in recent years have lessened this problem by combining more robust periodic estimations of orientation derived from the accelerometers and magnetometer readings. In cases where a tri-axial magnetometer is not included in the IMU then measures of orientation are influenced by heading drift around the gravity vector. Even in cases where a magnetometer is included, significant care must be taken to ensure proper magnetometer calibration and the avoidance of magnetic interference to achieve reliable measures of orientation.

As far as the author is aware, the first example of a single IMU was the 2010 study by Parnandi et at [62]. The system developed by Parnandi et al was composed of a single IMU placed on the wrist to capture movements. A single IMU system has also been implemented in the 2012 study by Zhang et al [63]. A clear limitation of only instrumenting a single segment is that it provides limited information if movements require the use of multiple segments in isolation or synergistically.

Subsequent studies have utilised IMUs placed on the torso, upper arm, and lower arm [12, 64, 65]. The additional information provided by instrumenting these segments has variously been used for the computation of joint angle, the ratio of energy between segments, and avatar simulation. One of the earliest examples of a multiple IMU study is in the 2011 study by Bento et al [12]. The study by Bento et al also implemented an additional IMU placed on the contralesional side to detect compensatory movements in the unaffected side. A follow-up study by Cruz et al [65] has also been performed using the same system. A diagram of the setup used in these two studies is shown in Figure 3.2. Finally, the study by Huang et al [64] placed an additional IMU on the hand which enabled the wrist joint angle to be computed.

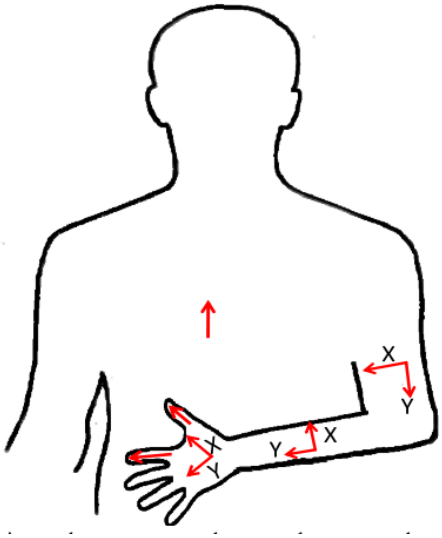


Figure 3.1: Wearable motion capture systems as designed by Patel et al [11]. Axes of single and dual-axis accelerometers represented by arrows

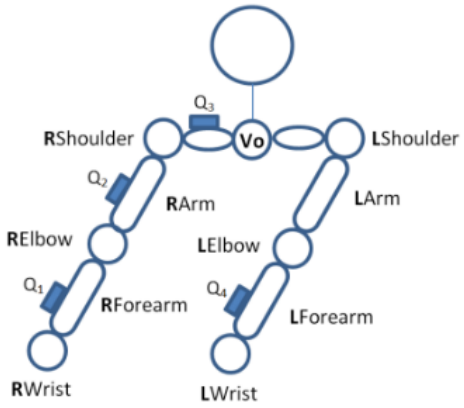


Figure 3.2: Wearable motion capture systems as designed by Bento et al [12]. IMUs are represented by blue rectangles

Visual Motion Capture

The Microsoft Kinect systems was the modality selected for visual motion capture in all prior automated systems. The main reason given for the selection was the low cost of the sensor, compared to lab-based motion capture systems, which makes it affordable for clinical implementation [13, 19, 66, 20, 67]. Another reason is that the Kinect has been proven to detect limb kinematics with reliable accuracy in a multitude of prior studies [68]. Lastly, the user-friendly Kinect Software Development Kit (SDK) provides access to the skeleton data, the video feed, and the microphone directly, which makes the system valuable for research applications [68].

There are several disadvantages associated with using the Kinect for clinical applications as compared to wearable sensing such as IMUs. The majority of these disadvantages stem from operation of limb tracking in the Kinect as compared to IMU systems. Whilst the IMU sensor is attached directly to the limb segment of interest, the Kinect system has to make approximations of the motion of limb segments based on particular landmarks. As a consequence, the Kinect system has difficulty tracking subtle movements as well as rotations around the bone axis. Secondly, the Kinect system is susceptible to occlusion, whereby the camera does not have vision of the limb of interest. This may lead to missing or noisy orientation data. One way in which the limitations of the Kinect system has been reported in the literature is as inaccuracies when tracking hand position as in the study by Otten et al [68].

One of the earliest Kinect-based automated systems was developed for the 2014 study by Olesh et al [13]. The study by Olesh et al used the Kinect to extract joint angles from the subject, which were then further derived into scores of motor function. An image of the setup of this study, with an illustration of tracked points detected by the Kinect system as compared to a high-end motion capture system, is given in Figure 3.3. A similar setup has been implemented in several subsequent studies [19, 66, 67].

A few recent studies have supplemented the data from the Kinect system with other wearable sensors [68, 69, 20]. This has been performed to combat the aforementioned limitations of Kinect, particularly the risk of occlusion and difficulty detecting hand movements. The study by Otten et al [68] paired the Kinect sensor with an IMU on the wrist and a wearable glove fitting with flexion sensors. The IMU enabled better detection of rotation around the forearm principal axis and tremor while the flexion sensors could detect grasp tasks. Similarly, the study by Julianjatsono et al [69] also supplemented the Kinect sensor with a glove. This glove was fitted with both an IMU and flex sensors. A final study by Lee et al [20] utilised a Force Sensing Resistor (FSR) to detect the grasp tasks of the FMA-UE. Overall, all these studies demonstrate the potential of fusion systems involving the Kinect sensor. One drawback of all the aforementioned studies is that they all required a device to be worn (glove). This method has drawbacks including fitting and hygiene concerns.

Robotic Motion Capture

Robotic systems offer some of the highest resolution of information possible of all the different systems discussed. Despite this advantage, these systems suffer from potentially constraining the subject during motion (leading to unnatural movements) or lacking the necessary Degrees of Freedom (DOF) to allow all required movements. The biggest limitation for clinical testing however is that these systems are expensive and often bulky meaning that they are not practical as a bedside clinical assessment device.

The study by Chongyang et al [70] is one of the earliest studies to assess the application of robotic devices in the clinical assessment of motor function. The combination of a hand-wrist rehabilitation and single-joint rehabilitation robot was used to assist or directly record metrics of motor performance. The main limitation of this study was that the robots were very limited in range of motion and therefore most movements of the clinical scale had to be detected using visual motion capture instead. This limitation was rectified in a subsequent study by Balasubramanian et al [71]. For this study, a specially

developed 5 DOF wearable exoskeleton was designed to enable measurement of a variety of different movements. This robot could measure joint angles as well as the forces output. The first robotic capture system to be implemented in a large clinical trial is the study by Bosecker et al [14] who tested their system on 111 chronic stroke subjects. This study used the commercially available shoulder-and-elbow InMotion2 robot (Interactive Motion Technologies, Inc.) to derive a variety of kinematic metrics from the subjects. One limitation of the study was that it was not possible to perform all the tasks from the clinical rating scales with this robotic system. Instead, a combination of custom and clinical rating tasks were performed, and measures derived from these movements. An image of the robotic system used in this study is shown in Figure 3.4.

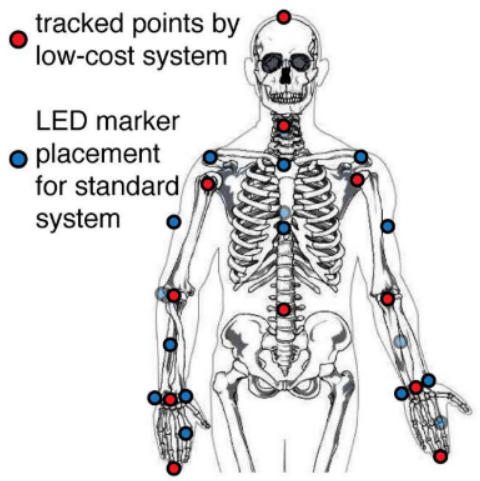


Figure 3.3: Visual motion capture system implemented by Olesh et al [13]. The low-cost system represents landmarks detected by the Kinect sensor

Figure 3.4: The shoulder-and-elbow InMotion2 robotic system as used in the study by Bosecker et al [14]

Myographic Capture

The introduction of MMG in a system of post-stroke motor dysfunction is only justifiable if it can be shown that there are differences in muscle activity post-stroke and that these can be meaningfully measured. Support for this is provided by a growing body of myographic research studies which have indeed detected significant changes between stroke and healthy subjects. In addition, three prior wearable system studies have already introduced EMG into their wearable system and have shown improved performance as compared to inertial sensing alone.

There has been a significant body of research already conducted to investigate the presence of statistically significant differences in parameters of EMG post-stroke as compared to healthy subjects. A study by Subramaniam et al [72] found significantly longer EMG burst duration (total duration of EMG activity) recorded from chronic subjects as compared to healthy subjects recorded during flexion and abduction tasks. It was postulated that this may be due to a reduced ability to produce optimal levels of neuromuscular activation post-stroke. Another finding has been a significantly different clustering index (CI) observed from paretic compared to healthy control subjects, as found in the studies by Tang et al [73] and Zhang et al [74]. The CI, as originally proposed by Uesugi et al [75], provides a measure of how "clustered" a signal is based upon the summation of differences in subsequent windows. Both a significant increase and decrease in CI was observed in the aforementioned studies, depending on which muscle was being investigated. An abnormal increase in CI is attributed to motor unit loss and changes in motor unit architecture whereas a reduction in CT is hypothesised to be due to muscle fibre atrophy [73].

As far as the author is aware, there has only been a single MMG study to investigate the differences in response between healthy and post-stroke subjects. The study by Hu et al [76] investigated subjects during isometric voluntary contractions of the biceps brachii and extracted the metrics of RMS and mean power frequencies (MPF) for analysis. At close to max voluntary contraction (MVC), the MPF and RMS values of MMG were found to be significantly lower for stroke as compared to healthy subjects. The explanation provided for this difference in response was an atrophy of the fast-twitch muscle fibres and a reduction of the neural input in the stroke-affected muscles [76].

There have thus far been three systems developed for quantifying motor function post-stroke to incorporate the myographic modality, and all of these systems have used EMG for this purpose. The study by Li et al [15] developed a combined IMU and EMG fusion system for the assessment of upper arm function post-stroke. This system utilised 11 EMG sensors and 2 IMU sensors positioned on the upper and lower arm, as displayed in Figure 3.5. Li et al found that the inclusion of EMG data, as opposed to IMU data alone, resulted in superior regression analysis of the clinical score. A study by Kim et al [77] used a simpler version of an IMU and EMG fusion system for the same application. This study used a MYO sensor (Thalmic Lab), which is composed of 8 EMGs and an IMU, placed on the lower arm. This was positioned to record kinematic data of this arm segment as well as myographic data at the brachioradialis muscle. Measurements were

made of the difference between tension and flexion of different muscles and this feature was found to be statistically significant between the impaired and less impaired arms. A final fusion system has been developed by Repnik et al [78]. This study used the MYO sensor, as used in the previous study, to record myographic data. Additionally, IMU devices were placed on the dorsal side of the hand, wrist, upper arm, and on the sternum. Myographic data was recorded to quantify grasp since this could not be captured using the attached IMU sensors. Muscle activity recorded using these sensors was found to correlate well with grasping activity as well as the level of grasping forces for stroke and healthy subjects.

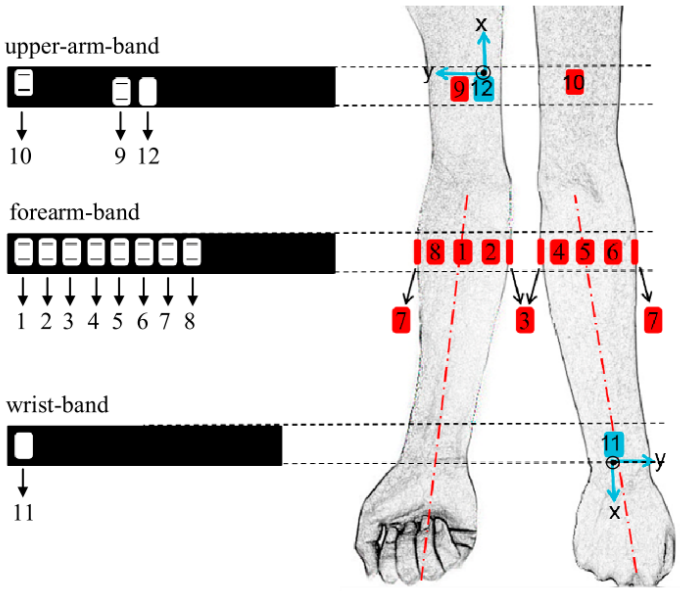


Figure 3.5: Placement of the sensors for the wearable system developed by Li et al [15]. EMG sensors represented in red and IMUs represented in blue

3.4.2 Data Collection

The majority of automated systems developed thus far have been included in pilot or proof of concept studies and have therefore been limited to small sample sizes (<15) of post-stroke subjects [13, 68, 79, 77, 59, 65, 20, 67, 12, 62, 71, 60, 63, 80, 81, 69, 82, 67, 72]. Another potential limitation in subject cohorts has been a reliance on healthy subjects to mimic levels of impairment [83]. As far as the author is aware, there have only been a handful of studies conducted with a large sample size (>= 30) of post-stroke subjects [84, 85, 19, 14] and only two of these have been with the wearable motion capture devices

of most relevance to the present study [85, 19].

The movements performed by the subject during testing must be selected to reveal motor impairment and preferably cover multiple components of upper-extremity dysfunction. For some studies, the choice was made to use an established clinical rating scale for this purpose including the FMA [13, 63], WMFT [17, 12], and ARAT [13]. An advantage of selecting tasks from existing rating scales is that these tasks have already been evidenced to be valid measures of motor function. A second advantage is that if the study involves the development of a predictor model, then this may be used to directly predict the clinical rating scale. This predicted score would therefore possess the validation of the clinical scale. Alternatively, some studies have instead chosen to use a custom-developed task such as simulated baseball swings [85] and custom reaching tasks [72]. Features derived from these tasks could then be correlated with standard clinical rating scales to prove their validity.

3.4.3 Data Segmentation

The subsection refers to the method by which the sensor data is windowed prior to the calculation of features. The window selected may have a significant effect on the usefulness of the resulting features. For instance, whilst most studies chose to window the data over the entire period of the task, some studies chose to further segment the task window by the subcomponents of the task. The study by Patel et al [16] chose to further segment the "lift can" task of the WMFT into the stages of "reach", "lift", and "drink". This fine segmentation approach was chosen because it was found that some stages of the task provided higher fidelity of information than others. By isolating these stages, the overall classification results were improved. An image of this segmentation is shown in Figure 3.6. The study by Zhang et al [63] went a step further by choosing a "fine-grained" approach. This involved segmenting each task into small fixed-time windows of 0.2 seconds each. Zhang et al found that this segmentation method enabled the capture of detailed patterns that standard clinical scales failed to reflect. Unfortunately, the study required hemiplegia of the subjects since the analysis required a direct comparison of an affected and non-affected side. Hemiplegia is not always present in stroke subjects and in cases where it is the "unaffected" arm is rarely truly unaffected. For this reason, the system presented in this study has limited applications.

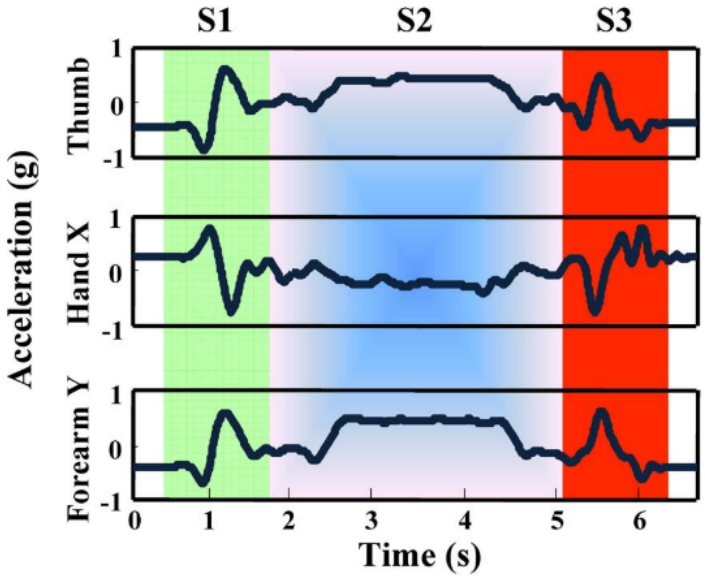


Figure 3.6: Raw accelerometer signal extracted while the subject performs the "lift can" task of the WMFT. Data segmented into reaching (red), lifting and drinking (blue), and return (red) stages. Taken from the study by Patel et al [16]

3.4.4 Feature Calculation

Studies which only instrumented a single limb segment with accelerometers or IMUs were limited to features for that limb and could not calculate composite features across limb segments. Despite this limitation, multiple sophisticated non-linear and frequency domain features were calculated for these studies as well as the more fundamental features. Some of the simpler sensor features which were calculated across most of the studies include the maximum and minimum, mean, standard deviation, root mean square, energy, and smoothness. More complex non-linear single sensor features calculated include the approximate entropy [59, 17], range of autocovariance [60, 17], and jerk metric [17]. Finally, frequency domain features calculated include dominant frequency [60] and the ratio of energy around the dominant frequency compared to the total energy of the signal [17]

Studies which utilised multiple IMUs or visual motion capture systems were able to compute more sophisticated measurements across different segments such as segment ratios and orientation. One feature based on segment ratios calculated was the intensity ratio of the accelerometers/ gyroscopes recorded across two IMUs [15]. Orientation features are a useful means of quantifying complex movements and are generally easier to interpret than more opaque features derived from the acceleration or gyroscope. Orientation feature derived in prior systems include joint angles [67, 20], segment rotation [20],

balance [64], and trajectory accuracy [64].

The have been a selection of myographic features extracted from the prior myographic kinematic fusion studies. The study by Kim et al [77] calculated a feature of muscle stiffness by finding the difference in tension of flexor with the tension of extensor at the brachial muscle. Muscle tension was estimated by standard EMG signal processing (rectification, amplification, and low-pass filtering). The fusion study by Li et al [15] calculated an EMG power distribution for the array of EMGs sampled. The Root Mean Square (RMS) of each EMG channel was first calculated. A percentage was then assigned to each EMG channel based on the proportion of the EMG value compared to the summation of the entire array. A similar method was also utilised for calculating muscle activity in the study by Repnik et al [78].

3.4.5 Feature Selection and Dimensionality Reduction

Reducing the feature set is an important component of any machine learning study in which the number of features exceeds the number of instances. The two main methods of reducing the feature space are by feature selection, whereby algorithmic methods are used to select existing features, and dimensionality reduction, whereby the feature matrix is transformed to a new reduced feature space.

The two forms of feature selection, filter and wrapper methods have both been performed in prior automated studies. Filter methods that have been implemented include the ReliefF algorithm [11, 17, 61] and L1-norm minimisation learning. A wrapper method that has been implemented is the use of SVR [64] and LASSO [15] for feature selection.

The study by Kim et al [19] utilised PCA as a means of reducing the feature space before feeding the data into a neural net for classification. The system developed by Bosecker et al [14] also utilised PCA prior to developing regression models of clinical score. An alternative application of PCA has been as a scoring method as used in two prior studies [13, 15]. This method will be discussed in more depth in the following section.

3.4.6 Feature Set Size

A secondary and oft overlooked consideration in feature selection is the size of the resulting reduced feature set. This has great importance since the size of the feature set has a significant effect on the training time and resulting classification performance. There

may also be a different optimal size of feature set depending on the task being classified.

There have been three main approaches used to determine the optimal feature set size. Firstly, there have been studies which have used an arbitrary size, or the same size as used in the previous studies. Secondly, there have been studies which optimised the feature set size for one or across several tasks and then used this number of features for all subsequent feature reduction (fixed feature set). Finally, there have been tasks which dynamically found the optimal feature set size for each task classified (dynamic feature set).

For prior systems which have implemented a fixed feature set size, the most common method of determining an "optimal" set size was by calculating or visualising the best feature set size across a subset of the total tasks. The number which best suited most of these tasks was then implemented as the fixed feature set size for predicted of all tasks. This method has been implemented in several prior automated studies [61, 17, 16, 84]. A plot of this as performed in the study by Del Din et al [17] is shown in Figure 3.7. Del Din et al found that RMS error did not decrease significantly across tasks when using more than 20 features so selected 20 as the fixed feature set size for subsequent feature selection. One drawback of fixed feature set methods is that they make the flawed assumption that the optimal feature set size for one task will also suit another task. This is likely incorrect for most clinical rating scales since tasks may be drastically different in terms of the type of movement captured or limb segments involved.

An improved way of implementing feature selection is to dynamically find the optimal feature set size for each task being predicted. As far as the author is aware this has only been implemented in the prior automated study by Kim et al [19]. In this case, PCA was used to reduce an original feature set of 100 down to between 4 and 10 dimensions for each assessment task.

3.4.7 Building the Predictive Model

The studies discussed thus far may be categorised depending on whether they sought to find a correlation with or a prediction of the clinical rating scales. The focus of this section is on the development of predictive models and as such correlation studies will be excluded from this section. Prior predictive models have been developed which have derived a score via classification, regression, or by a novel scoring method. A distinction is made in this section for studies which used "off-the-shelf" classification models which may be trained for a variety of functions, as opposed to custom classification models

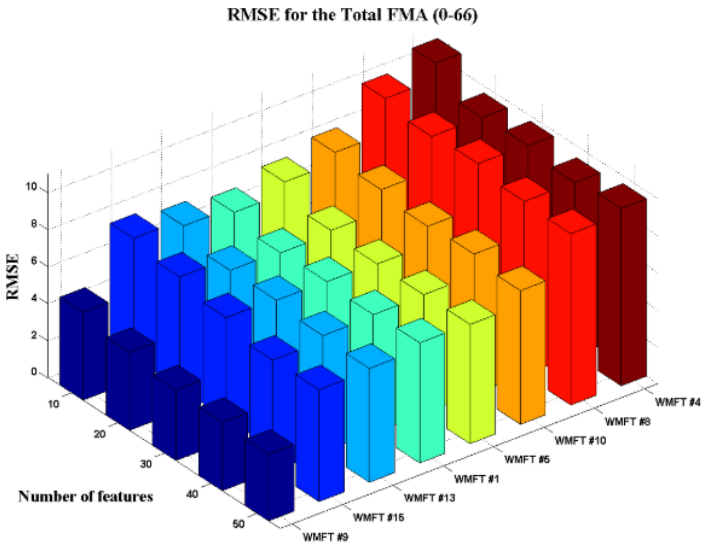


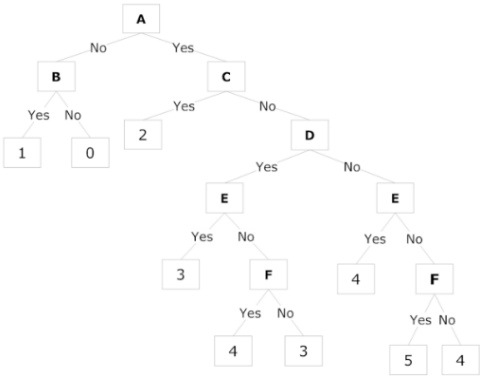
Figure 3.7: Graph of the effect of different feature numbers used for predicting the scores of different tasks of the WMFT. Taken from the study by Del Din et al [17]

which have been created using domain knowledge specifically for the study. These will be referred to as generalised and custom classification models respectively.

Several different generalised classification models have been implemented in prior automated systems. These include linear SVM [68, 83], ANNs [68, 19], and logistic regression [79]. In addition, the ensemble classifier random forests has been applied [11, 17]. It is difficult to draw any kind of direct comparison between the classification results of these studies due to widely varying parameters including subject pool size, instrumentation, and clinical scale evaluated. Ensemble classification methods, such as random forest, utilise the aggregate result from muscle classification models and have been found to perform better for small datasets. A common link for all these classification models is that they are reliant on sufficient training data to find the optimal coefficients for classification. Many of the stated studies utilised relatively small sample sizes and as a result, the classification models would be unlikely to generalise well to new data.

Custom classifier models have thus far been implemented in the form of decision trees [65, 12, 67] and a binary logic classifier [20]. A diagram of the custom decision tree developed for the study by Bento et al [12] is illustrated in Figure 3.8. Custom classification models are well suited for clinical rating scales in which the definition of score is based on specific movement parameters since these parameters may be easily integrated into the classification model. These systems would also be expected to perform better than generalised classification models for small datasets since they are not dependent on a large amount of training data to achieve the optimal parameters. One limitation of

this method is that it relies on human judgement, in the form of domain knowledge, to determine optimal features and parameters. This means that there may be useful features for classifying tasks which are missed because they are beyond the comprehension of the model designer



DECISION TREE FEATURES

Feature	Description
A	Task Completed
B	Detected movement in involved arm
C	Detected movement in uninvolved extremity
D	Movement synergy with the shoulder joint
E	Detected movement out of the plane of action
F	Is the movement smooth

Figure 3.8: Custom decision tree for classifying scores from the WMFT, developed by Bento et al [12]

Regression models have been used to determine a continuous rather than discrete output of clinical score. Methods implemented include LASSO [15], a form of linear regression which uses shrinkage to shrink data towards a central point. Other methods of regression have included linear [60, 14, 69] and ANN [69] based regression models. An oft stated advantage of these studies is that regression holds the potential to raise the resolution of the output score. However, the accuracy of the score is still limited by accuracy of the clinical rating scales which form the labels. Since the clinical rating scales are based on discrete, ordinal scores, there is unlikely to be any additional accuracy in the use of regression over classification for this application.

Finally, there have been novel scoring methods whereby the scoring was achieved independently of a clinical rating score. Thus far these methods have operated based on a comparison between the affected and unaffected arms post-stroke. In the study by Zhang et al [63], the same tasks were performed first with the affected and then the unaffected arms. Dynamic time warping was then used to enable direct comparison of the features calculated for each arm. Finally, a similarity metric was calculated to determine the similarity of performance between the two arms, with a low score indicating impaired performance of the affected arm. In a study by Olesh et al [13], PCA was used to deconstruct the principal components of averaged healthy joint angles. The same principal components were then used to construct the temporal joint angle profiles of

the affected arm. A coefficient of determination was then used to calculate a score based on how similar the reconstructed affected joint angles were to the original affected joint angles. This score shows how well the principal components of the unaffected arm represented the movement of the paretic arm. An advantage of both these studies is that they determine a score which is independent of the clinical rating score. This means that the inherent limitations of the clinical rating score are not integrated into the model. This can however also be considered a limitation since the score produced does not possess the validation associated with the clinical rating scales. Another limitation is that both studies draw comparisons based on the affected and unaffected arms of the same subject and therefore require the presence of hemiplegia. This is not always the case in post-stroke populations and therefore these systems have limited clinical application.

3.4.8 Review

A final review is made of all the aforementioned automated studies with particular emphasis made on any novel implementations or limitations which should be taken into consideration in the present study. This is discussed with respect to instrumentation, sample size, feature extraction, and predictive models.

The prevalent instrumentation utilised in early automated systems has been a solely accelerometer-based setup. This sensor system has low energy requirements which makes it well suited for pervasive monitoring systems but is less well suited as a clinical quantification system due to the limited motion information provided. The use of IMUs, particular in multiples, and the Kinect sensor offers much greater fidelity of information which in turn may improve prediction performance. Additional information that may be derived includes joint angles which form important parameters in many clinical rating scales [86]. The Kinect has the advantage over an IMU-based system by being the more established and well-validated system for the measurement of joint angles. The Kinect also possesses a low cost and a simple interface for extracting clinical data. Disadvantages of the Kinect sensor include high set up time, risk of occlusion, and poor performance when quantifying rotation around the principal joint axis. The final two disadvantages mean that the Kinect should be supplemented with a secondary device to provide sufficient results. When testing a subject population with a wide range of motor deficits (including bed-bound subjects) multiple IMU systems are likely the superior sensor modality due to their robustness, lack of occlusion problems, and easy setup and use in any environment. The Kinect sensor is less practical as a bedside assessment mea-

sure due to the requirement of cameras being set up and the subject being in an upright position. For applications where subjects are mobile enough to participate, the Kinect sensor may be the preferred sensor system due to its high validity and ease of use. Three recent studies have integrated myographic sensors into their sensor system as a means of supplementing the kinematic data with information about muscle activity. Early results are promising with all studies reporting that the inclusion of muscle activity provided higher performance than kinematic data alone. In addition, myographic sensors hold the advantage of being able to quantify motor areas which are otherwise difficult to measure kinematically (such as grasp function).

A small sample size is a persistent limitation in most prior automated studies. Testing on such a small sample size reduces the scientific significance of any findings and may provide misleading results. In addition, the majority of studies relied on classification or regression models to make predictions. A small training set would mean that these models would be expected to perform less well and would also have to be training on low-dimensional feature sets to avoid the risk of over-fitting. The presence of small sample sizes in prior studies is likely due to difficulties encountered recruiting suitable post-stroke subjects. Now that the field of automated evaluation of motor-function post-stroke has become more established, there is a requirement for more follow up studies on larger cohort sizes to better validate the achieved results.

Numerous features have been calculated across automated systems and these are rarely consistent between studies. This makes evaluation of the most useful feature subset difficult. Accelerometer and gyroscope metrics alone provide a basis for deriving a multitude of different features. These features provide useful motion information but are largely dependent on how the subject chooses to perform the movements. If the subject elects to perform tasks quickly then these movements will possess more energy, and this may provide misleading results. An exception to this is jerk, calculated from the differential of acceleration, which can be used to calculate the jerk metric, a useful metric of tremor [87]. The jerk metric is normalised using the velocity of the movement and therefore is not dependent on the speed the subject elects to perform the action. Orientation-based features are likely to be more useful for classifying clinical score than accelerometer or gyroscope-based features since they correspond more closely to clinically relevant parameters such as range of motion, balance, and segment rotation. These features are also more transparent (correspond to known metrics such as joint angle) and as such may be more easily understood by both the subject and clinician. Finally, two

features have so far been derived from EMG data. These have been based on the tension between different muscles and a power distribution derived from an array of EMG sensors. Both these features have been shown to provide a useful insight into muscle activity.

Most automated systems applied classification models to predict clinical score. This method appears preferable to regression models since the labels used for training are discrete and although ordinal, are not linear. This means that a regression model is unlikely to provide any increased accuracy as compared to classification. Custom classification models were developed which utilised custom features and thresholds based on existing clinical rating scales and domain knowledge. These models worked well on small sample sizes which would have provided insufficient training data to train more complex models. Lastly, two novel studies developed scores based on comparative metrics based on the affected and unaffected arms. This resulted in sophisticated scores which go beyond that offered by the standard clinical rating scales. A major limitation of these studies is that they are not suitable for most clinical studies due to the requirement for hemiplegia in the cohort of subjects. This condition is not always present post-stroke and when present will occur at varying degrees of severity.

In summary, it is the author's opinion that the optimal sensor system for bedside automated clinical evaluation be composed of multiple IMUs and supplemented by myographic sensors. Multiple IMUs would be easy to attach to subjects at all impairment levels, while the myographic sensing would solve the problem of quantifying the hand and wrist motor function involved within grasp or manipulation tasks. There have not yet been any studies which have offered a definitive feature set for such automated systems, but orientation-based features relate most strongly to existing clinical parameters and would likely be useful for this application. Finally, novel methods of deriving a score (independent of the clinical rating score) are promising but are currently dependent on comparisons between affected and non-affected arms which is a flawed methodology. Typical methods have involved "off-the-shelf" or custom classification models and both these methods should be investigated further.

3.5 Instrumentation

The instrumentation utilised in the present study was composed of multiple in-house data logger boards (each housing a single IMU) and MMG sensors. These two sensor types will be discussed within this section in terms of their function, design, configuration, and synchronisation.

All of the instrumentation discussed in this section was developed in the Biomechatronics lab (Imperial College London). The data logger boards were produced prior to the start of this PhD project and simply applied in the present study. The MMG sensors were built as part of the project using a design which has previously been established. All of the cases and straps used for the devices were designed and printed as part of this PhD project.

3.5.1 Data Logger Board

The data logger board utilised for the present study has been developed in-house (Biomechatronics Lab, Imperial College London). A diagram of this board is illustrated in Figure 3.9 and highlighted with several key components as listed below:

1. Microcontroller (PIC24FJ64GA104)
2. Wireless Module (BT900)
3. IMU (LSM9DS1)
4. Micro-USB port
5. 8 broken-out Analogue-to-digital converter (ADC) pins for sampling data from external devices

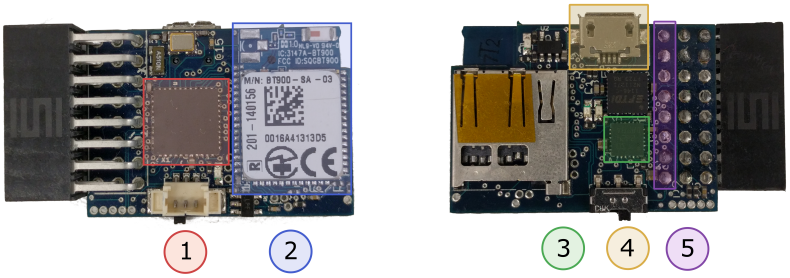


Figure 3.9: Key components of the data logger board

The primary microcontroller of the data logger board is a 16-bit microcontroller (Microchip) with a processor speed of 16 Million Instructions per Second (MIPS). This microcontroller contains several useful peripheral modules including clock, I^2C modules, Universal Asynchronous Receiver Transmitter (UART) modules, and external interrupts. These modules were all necessary for recording from the IMU chip and transmission of data over Bluetooth. Another useful property of the microcontroller is that it enables up to 8 pins to be purposed as ADC pins (see subsequent paragraph on broken-out ADC pins).

The BT900 wireless module provides dual-mode implementation of Bluetooth version 4.0. This provides a low energy means of Bluetooth communication at a suitable communication range and bandwidth for the present study. Bluetooth streaming directly to a master device such as a computer is a useful property in research devices for online visualisation or data logging for instance.

The LSM9DS1 chip was fitted on board for the measurement of inertial data. This IMU contains triaxial measurement of accelerometer, gyroscope, and magnetometer. This provides 9 DOFs in total which is optimal for data collection and subsequent orientation calculations. The miniaturised size and easy configuration of this chip makes it a straightforward method of implementing onboard wearable sensing.

A micro-USB port is included in the data logger board for charging the battery and for wired serial communication over USB (as an alternative to Bluetooth communication for data collection).

The maximum configurable number (for the microcontroller) of 8 ADC pins were utilised on the data logger board. These pins were broken-out on the boards for this project by using a 3 x 8 set of female 2.54 mm pitch header pins. These external header pins enabled the sampling of data from external devices, such as the MMG sensors used for this study.

3.5.2 Mechanomyogram

The MMG used for this study was developed in-house (Biomechatronics Lab, Imperial College London). The sensor uses a glue-less design, the structure of which is composed of three polymer (3D-printed) parts: the housing, sleeve, and clip. The actual microphone is contained on a small printed circuit board (PCB) and positioned within the housing. The housing contains a chamber which has been designed to achieve the highest gain of the MMG signal while maintaining the flattest frequency response [88]. Finally, a mylar

membrane is fitted over the chamber using the sleeve to create a pressurised chamber. A diagram of the aforementioned components of the MMG, as well as a photograph, are displayed in Figure 3.10 and 3.11 respectively.

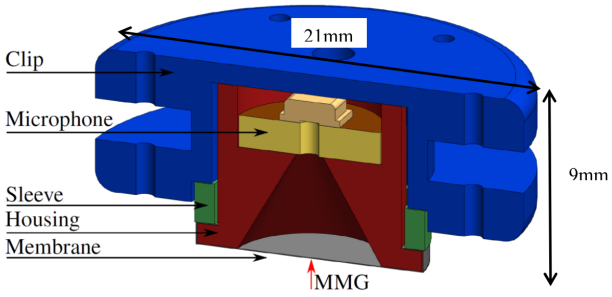


Figure 3.10: Computer model of the in-house mechanomyogram model including the key components for construction (Biomechatronics Lab, Imperial College London)

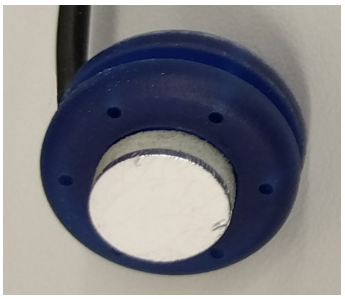


Figure 3.11: Photograph of the in-house mechanomyogram (Biomechatronics Lab, Imperial College London)

The signal that is collected from the microphone of the MMG represents the change in internal pressure of the chamber caused by the distortion of the membrane. When placed flat against the surface of the skin above a muscle, the signal is analogous to the change in pressure caused by muscle vibrations which have propagated through the skin and mylar membrane. This vibration provides information which relates to an intrinsic property of muscular contraction [49].

3.5.3 Sensor Configuration

The testing phase of the study required the instrumentation of different body segments. These sensors would then transmit data while the subject performed a series of pre-prescribed tasks. For some body segments only inertial data was gathered, which required only the data logger board to be attached to the segment of interest. For other body segments both inertial and myographic data was to be gathered. In this case, the MMG sensors were wired directly to the broken-out ADC pins of the data logger board and sampled appropriately.

All cases and straps for the data logger boards and MMGs were specially printed for this study using the Form 2 3D printer (Formlabs). This is a stereolithography (SLA) type of printer which uses a laser to harden a UV-sensitive layer. This printer was selected

since it provides a high-resolution (minimum layer height of 25 microns) of printing in a small form size and offers a wide selection of different resins including the flexible type (as detailed later).

The MMG and data logger board cases were printed using the tough resin. This resin was selected since it provided good level of strength (ultimate tensile strength of 55.7 MPa [89]) in the finished part and could be printed at a high resolution specification. A high strength case was required to ensure good durability and protection to the electronic parts, while the high resolution was desirable to ensure attachment parts on the case would function well. Images of the final printed cases for the data logger board are shown in Figure 3.12 and Figure 3.13 respectively. A special case had to be designed for the data logger board positioned on the torso since this could not be attached to a limb segment (like the other cases). This case was printed to incorporate clip-on buttons which could easily be attached to an elastic strap.

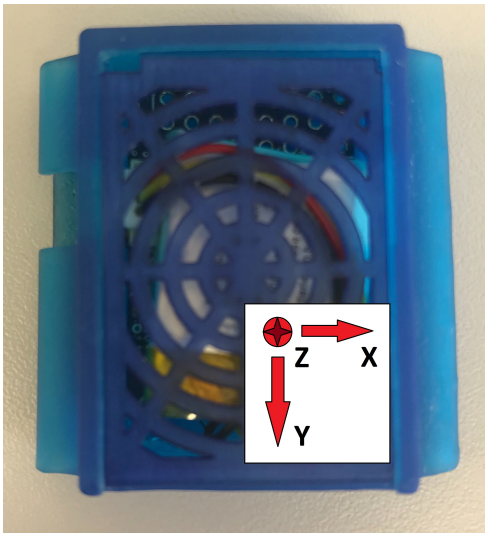


Figure 3.12: Top-down view of the data logger 3D printed case. Local axis frame for the accelerometer/ gyroscope included where the z-axis points upwards

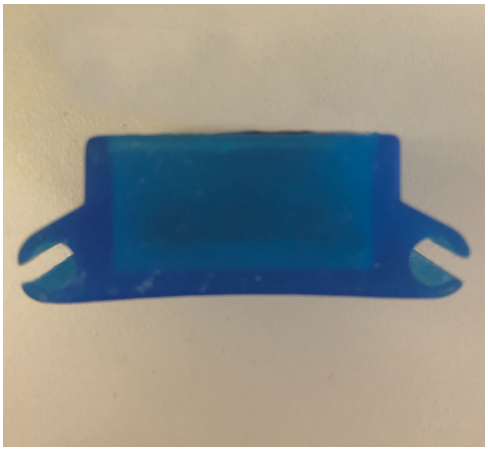


Figure 3.13: Front view of the data logger 3D printed case. "Wings" visible at the sides of the case enable fitting of the wearable straps

Wearable straps, for the data logger and MMG cases positioned on the limb segments, were printed with the flexible resin. This resin type was selected since it provided a comfortable fit but was less prone to failure than the elastic type of resin. A range of sizes of wearables straps were printed to ensure that the sensor system could be fitted to a wide range of limb sizes. Wearable straps could be easily switched out prior to clinical testing to ensure the best possible fit and in turn minimise subject discomfort and sensor

migration during testing. Velcro was applied to the wearable straps to enable them to be affixed to the subject limb segments. Images of the straps attached to the data logger board and MMG sensors are shown in Figure 3.14 and Figure 3.15 respectively.

Figure 3.14: Top-down view of the data logger with flexible polymer straps attached

Figure 3.15: Top-down view of two of the MMG sensors with flexible polymer straps attached

3.5.4 Sensor Synchronisation

The sensor setup implemented in the present study is composed of multiple sensors across two different modalities (inertial and myographic). This complexity meant that sensor synchronisation was a primary consideration for this study. In particular, there were two issues which had to be addressed in this domain. Firstly, the two different sensing modalities were recorded from different devices and at different frequencies. Secondly, the data logger boards were not all wired together and as a result clock synchronisation was a major concern.

The synchronisation of the MMG sensors was achieved by sampling these sensors directly from a master data logger board. This meant that the MMG signal was collected in analogue form and then digitised within the data logger board itself. This method ensured that the same clock would be used both for sampling the MMG sensors and the on-board IMU. Although the frequency rates differed for the two modalities, the use of

the same clock meant that these were easily synchronised in post-processing. For the present study MMG sensors were located on the forearm and as such the data logger board located on the wrist was selected to sample these sensors. Additional data logger boards were located on the upper arm and torso, as explained in more detail in Subsection 3.7.2: Instrumenting the Subject

Synchronisation between the different IMUs was more complex since these devices each had their own clock. This was handled by generating a packet at the data logger board which included the device ID and the packet number. This enabled any dropped packets to be detected at the computer-side and provided the exact sampling rates of each device so that these could be synchronised in post-processing. Finally, each packet received at the computer-side was assigned a time stamp so that any periods of extended Bluetooth drop-out could be detected, and data could be synchronised between devices. An image of the data synchronisation steps is shown in Figure 3.16

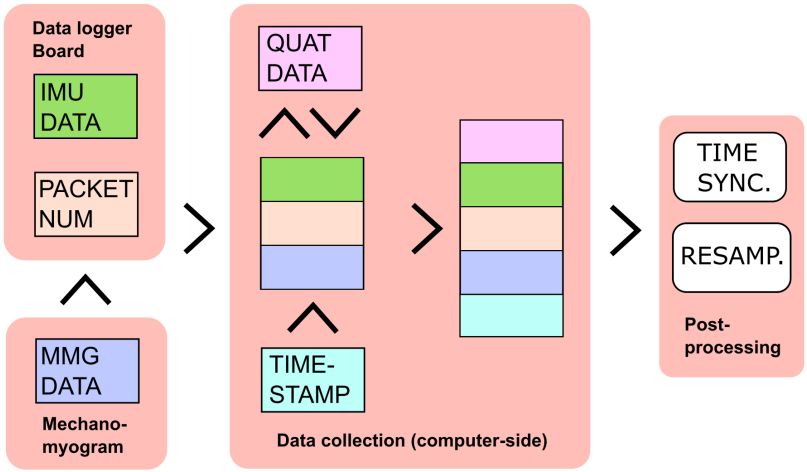


Figure 3.16: Project data synchronisation and output. Packet number (provided by the data logger board) and timestamp (provided at the computer-side) helped ensure the data was sampled correctly and synchronised between data logger boards

3.6 Software

Multiple custom Graphical User Interfaces (GUIs) were developed to perform the stages of data collection, processing, and classification. Particular focus was made to ensure the GUIs were designed to meet the goals of the project while also providing enough

flexibility to be further adapted for future work. The designed software will be discussed in terms of their primary functions, programming language, and key libraries. Software is categorised depending on whether it was involved in the online or offline data collection/ processing stages of the study.

All of the software documented in this section was developed in the Biomechatronics lab (Imperial College London) by the author specifically for this PhD project. Software was developed using either the Python or C# programming language using the Qt (The Qt Company) and .NET (Microsoft) frameworks respectively.

3.6.1 Online

The first major piece of software developed for this study was the "Data Saver" GUI. The primary function of this GUI was for forming Bluetooth connections with and streaming data from the data logger boards. The GUI also provides a variety of functions to transform the raw transmitted data into a usable format. Finally, the GUI provides options to visualise the transformed data or save the data for later processing (see "Offline" section).

The C# programming language and the .NET framework were selected for the GUI development. The C# language was chosen due to the ease of generating simple and intuitive user interfaces with the .NET framework, and for its speed for multi-threaded applications. The latter property was particularly important for the present study since multiple processes were performed simultaneously including displaying the GUI, maintaining up to three Bluetooth connections, visualising data, and saving the data. A number of additional libraries were implemented which are not included in the standard .NET framework. These were the *InTheHand* library for forming Bluetooth connections, *OxyPlot* for real-time graph visualisation, and *OpenTK* for real-time avatar and object visualisation. An image of the connection tab of the Data Saver GUI is shown in Figure 3.17.

The primary function of the online GUI was for directly logging data from the post-stroke subject and as such it was designed to be used by someone without training in software development. The idea was that this could be adopted directly by clinicians assessing the subject or even by a carer for home-based assessment. For this reason, the GUI was developed with a focus on usability, with large, labelled buttons and an intuitive design.

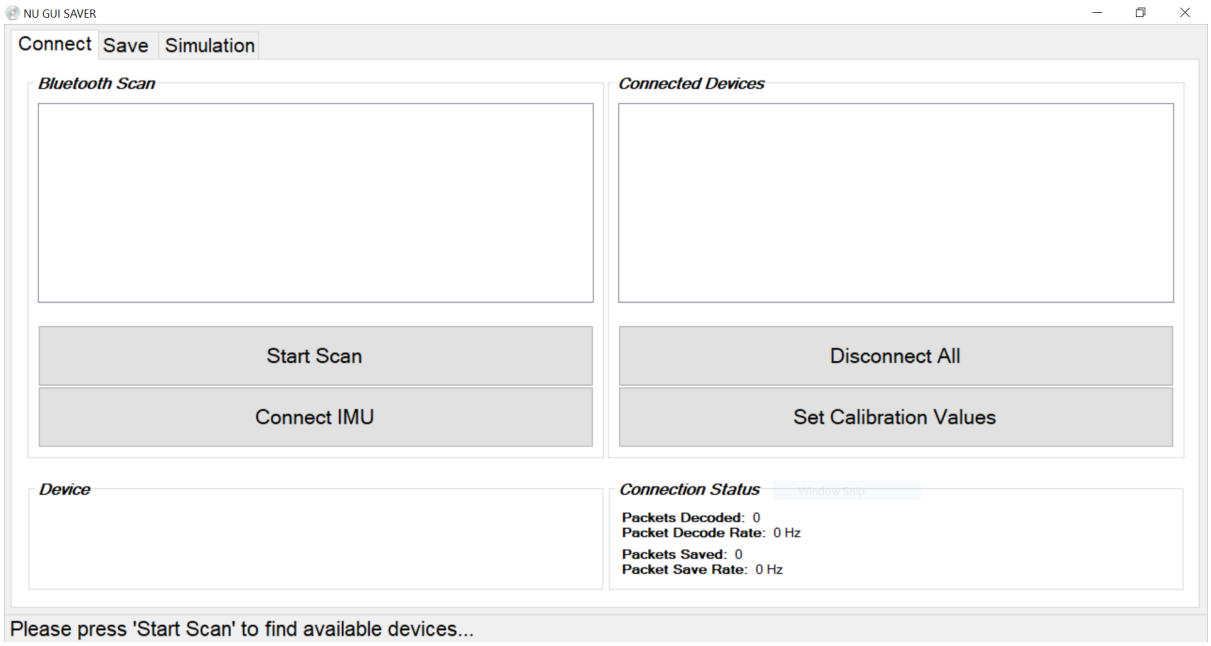


Figure 3.17: Data Saver GUI- Connection tab. Buttons provided for performing a scan to show all available Bluetooth devices, forming Bluetooth connections, and setting calibration values for the device

3.6.2 Offline

This subsection covers all the pieces of software which were developed for offline data processing in this study. All software in this section was developed using the Qt framework and implemented in the Python language. The Qt framework was chosen for its user friendly and cross-platform GUI development. The Python language was selected due to the wealth of libraries it has available both for managing large datasets and visualising them. A selection of python libraries used in all the python-developed GUIs were *Pandas* for large DataFrames, and *Matplotlib* and *Seaborn* for graph plotting.

The offline GUIs were developed with a focus on providing a wide range of functionality and to be adaptable enough to modify for use in similar studies. An object orientation approach to design was taken to allow for easy maintenance and modification of the code base. The use of the Qt framework meant that the GUI could be operated in a wide range of operating systems. Even though a visual rather than script-based interface was chosen for this software, a programming background is recommended for operation of these GUIs due to the requirement to ensure that any data files imported follow the correct formatting. For this reason, these GUIs are targetted for use by someone with a software background and may not be suitable for use by a clinician without significant training.

Data Processor GUI

The "Data Processor" GUI was developed for data processing, plotting, and feature extraction. The additional library *SciPy* was used for filtering and feature calculation, and the library *SQLAlchemy* was used to utilise a mySQL database for data manipulation when there was insufficient local memory. The import data tab of the Data Processor GUI is shown in Figure 3.18.

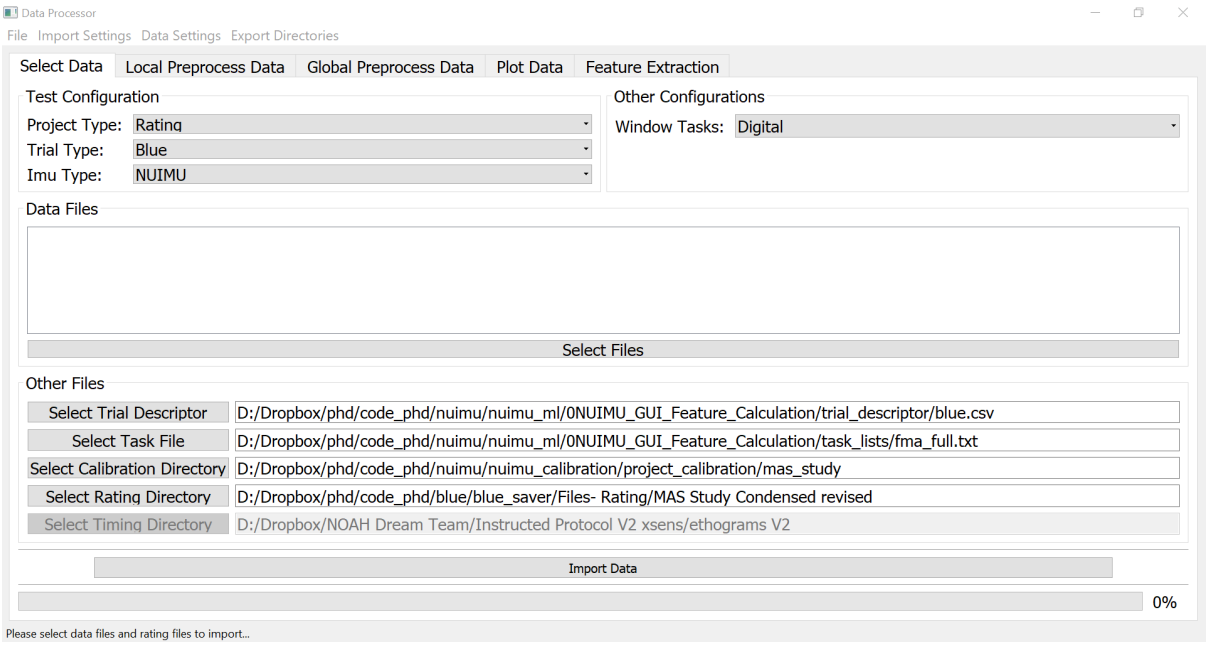


Figure 3.18: Data Processor GUI- Import Tab. Buttons provided for importing trial sensor data as well as calibration, rating, and task files

Feature Explorer GUI

The "Feature Explorer" GUI was developed for the visualisation and manual removal of features extracted from the data. Functions of this GUI include the plotting of features against each other or an independent variable to study their usefulness and the inter-correlation between features. This GUI also provides functions to enable straightforward manual dropping of features from the data. This is a requirement for this study since some of the tasks performed are only captured by a subset of features.

Feature Classifier GUI

The "Feature Classifier" GUI is the final piece of software developed for this study. This GUI enabled the creation of custom classification pipelines to generate and test the performance of classification models. This pipeline incorporates all the critical stages of classification testing including feature standardisation, balancing, selection, and classification. A wide variety of feature reductions method and classifiers were made available, and these different methods could be included as hyper-parameters in a single pipeline analysis. Lastly, both cross-validation and hold out methods of classification performance testing were provided.

Additional Python libraries utilised include *scikit-learn* for feature preparation, selection and classification, *imbalanced-learn* for data resampling and creating the classification pipeline, *TensorFlow* for the development of ANNs, and *lightGBM* for the implementation of the light gradient boosting classifier.

3.7 Data Collection

The section covers all the major components of data collection for this study including subject recruitment, instrumentation, and testing.

3.7.1 Recruitment

A cohort of 64 subjects was recruited from the acute and hyper-acute wards of Charing Cross Hospital (London). There were no subject requirements in terms of the minimum level of required motor function (unlike many prior comparable studies). This meant that subjects could be included in the study even if they were bed-bound. Cognitive ability was assessed before testing by the Glasgow Coma Scale (GCS). Subjects were excluded if they did not achieve a full score to ensure the subject was capable of giving informed consent to participate in the study and was able to follow verbal commands. Finally, a questionnaire was completed by the subject prior to assessment to determine if they were suffering from depression as this may affect their motivation to perform the given motor tasks. Subjects were excluded who scored below a certain threshold for this scale.

Full subject demographics are given in Appendix A Table 8.2. A summary of these demographics are as follows:

- Sex: 33 male, 31 female
- Affected Side: 43 left, 21 right
- Age: range 33-95, mean 66.3, SD 13.8
- FMA-UE Score: range 0-60, mean 42.7, SD 18.1

3.7.2 Instrumenting the Subject

This subsection details how the instrumentation was actually attached to the subject. Locations are presented with respect to the standard anatomical position for the sake of clarity.

The data logger boards were attached to the lower arm, upper arm, and torso of the subject. Each of these body segments was chosen due to their contribution to upper extremity tasks. In addition, derivation of the orientation at these segments enabled the calculation of joint angles at the elbow and shoulder. The arm to instrumented was selected based on which side of the participant's body demonstrated the most weakness. The data logger board for the lower arm was placed on the wrist due to the ease of fitting at this location. The upper arm was instrumented by placing the data logger board just above the elbow in position in-line with the lower arm data logger. The torso data logger board was positioned in the centre of the chest using a custom elastic strap. A diagram of the instrumented subject is shown in Figure 3.19. Fitting of the sensors was optimised by selecting the flexible straps which best suited the size of the limb segments of the subject. After testing, all sensor cases and straps were sanitised using alcohol wipes to avoid the risk of cross-contamination.

Two MMG sensors were placed on the mid-forearm and connected by wires to the data logger board attached to the wrist. The MMG sensors were located above the flexors of the fingers (flexor digitorum profundus and flexor digitorum superficialis) and the flexor carpi radialis. These two muscles groups have actions in finger and wrist flexion respectively. The recommended layouts for myographic recording from these two muscle groups are shown in Figure 3.20 and Figure 3.21. It is important to note that these diagrams were developed for instrumenting the muscle bellies via EMG rather than MMG. These diagrams were used as guides for the placement of the MMGs regardless due to a lack of comparable documentation available for MMG, and because the nature of MMG means that it records a larger aggregate of muscle activity and as such positioning

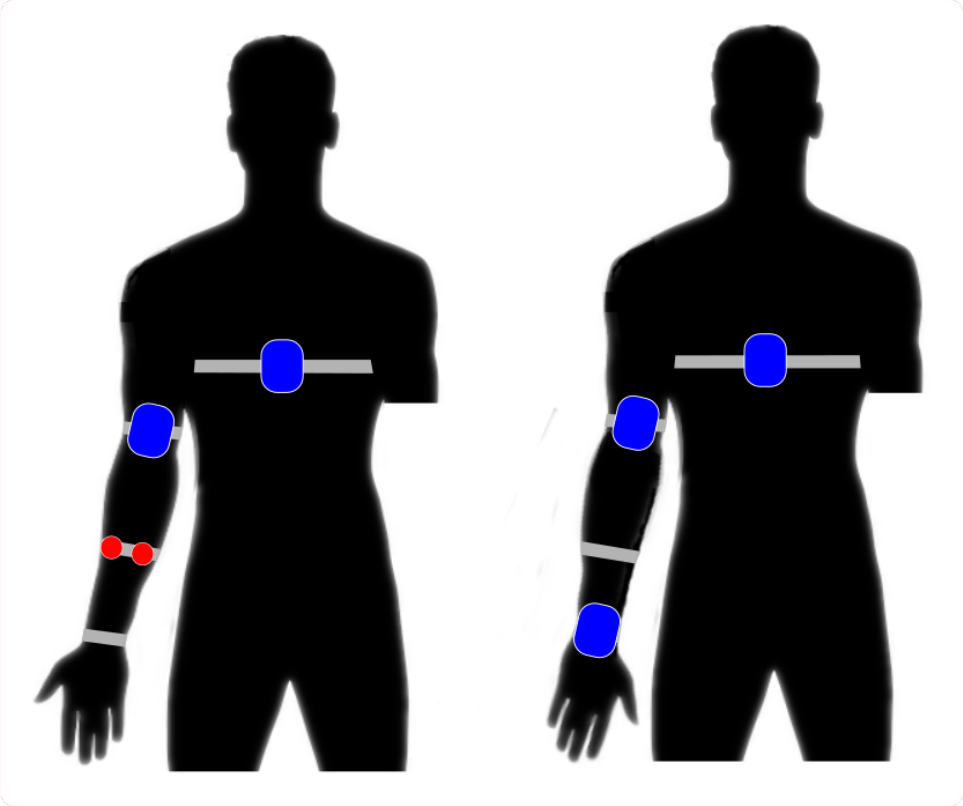


Figure 3.19: Diagram of the sensor setup up (assuming lesions on the right-hand side). Body in the standard anatomical position (left) and the same with the exception that the arm is pronated instead of supinated (right). Data logger boards are represented in blue and mechanomyographic sensors represented in red

of these sensors is considered to be less critical than for EMG. By instrumenting these muscles group it is possible to assess these motor actions they control, which are otherwise difficult to measure kinematically. An image of the MMG sensors attached to the subject forearm is shown in Figure 3.23. An image of the wrist data logger board (used for recording IMU data and sampling from the MMG sensors) attached to the subject is shown in Figure 3.22.

The two attached MMG sensors were expected to provide measures of flexion of the wrist and fingers. These are important measurands for both the wrist and hand subsection of the FMA-UE clinical rating scale that this study is attempting to instrument. It is worth noting that this study does not measure any information about the complementary extension muscle activity of the wrist and finger flexors. As a result, it is only possible to capture partial information relating to certain tasks, such as circumduction and repeated dorsi-volar flexion, and no information relating to the task of mass extension. This clearly presents a significant flaw in this study. The omission of this region was due to difficulties sampling from further MMG sensors with the data logger board. Any future work that

may be conducted with this sensor system should try to rectify this problem so that both the flexors and extensors may be fully captured.

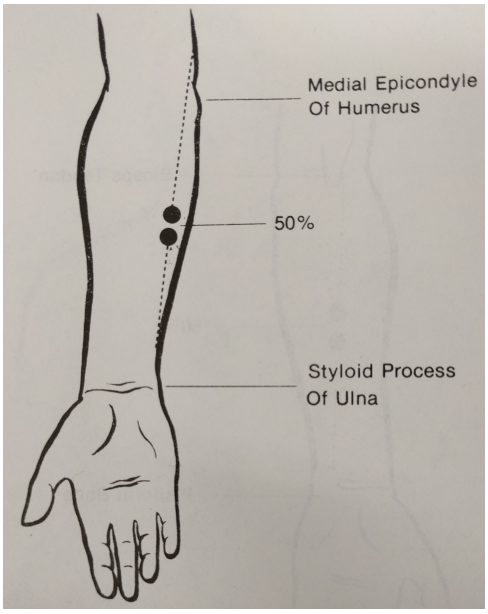


Figure 3.20: Recommended layout for myographic instrumentation of the finger flexors. Taken from [18]

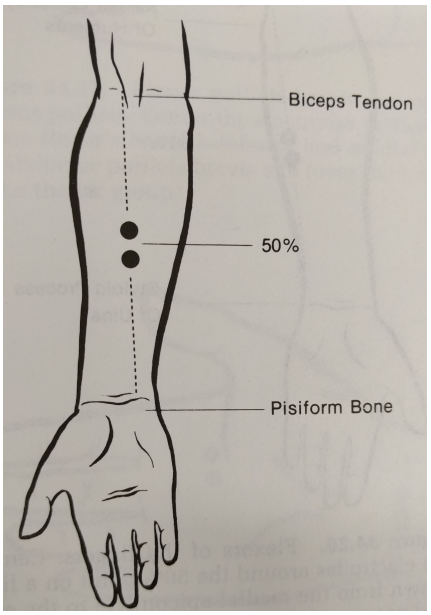


Figure 3.21: Recommended layout for myographic instrumentation of the flexor carpi radialis. Taken from [18]

Figure 3.22: Data logger board attached wrist of the subject

Figure 3.23: MMG sensors attached the forearm of the subject. Note that the sensors are slightly misaligned in this photo for recording of the finger flexors and flexor carpi radialis

3.7.3 Clinical Protocol

The clinical protocol for this study instructed subjects to perform a set of pre-prescribed tasks while wearing the novel instrumentation as discussed in the previous section. The task set selected was the upper extremity sub-section of the motor function section of the FMA. The FMA was chosen since it is one of the most widely used and well-validated clinical scores of motor function post-stroke [90]. In addition, the tasks included cover all body segments of the upper extremity including grasp and hand function. Movements of this scale are assessed both in isolation and synergistically to gain a detailed assessment of motor dysfunction. This provides a rich domain from which to collect sensor data and extract useful features.

One of the main differences conducting the FMA-UE for this study as compared to conventional clinical assessment was the additional setup time required for attaching and connecting to the sensors. Secondly, the subject was instructed to pause before and after each task to be able to digitally segment the task period. These two requisites meant that

instrumented testing required more time as compared to conventional clinical testing. Despite this, the additional time was lower than would be expected for camera-based setups (such as the Kinect) since this method would require subjects to be transported or a camera setup before testing. The total assessment time for this study rarely exceeded 20 minutes. This is considered acceptable for clinical research studies but may be too long for conventional clinical assessment as part of normal post-stroke care.

3.7.4 Clinical Assessors and Training

The nature of the clinical protocol for this study meant that two examiners were required to assess the subject. One examiner was in charge of applying the instrumentation to the subject and logging the data. The other examiner was responsible for performing the actual FMA, including interacting directly with the subject for the tasks which required this.

For the first batch of subjects tested ($\cong$ 30) a doctor with significant experience performing the FMA on the stroke ward was on hand to perform the clinical rating assessment. In this case, the author of this paper was responsible for setting up the instrumentation and logging the data. Due to the imminent departure of the doctor to another hospital, it was necessary to find an alternative clinical assessor. A student undertaking a masters in neuroscience was assigned to the project and was trained by the doctor to perform the FMA. This involved one to one training as well as being referred to a number of instructional online videos. There was also phasing in period whereby the experienced doctor was on hand to observe the performance of the student during assessment of the next few ($\cong$ 5) subjects to ensure consistency of scoring. Finally, to minimise discrepancies of scoring between each rater, an avatar visualisation of the tasks was also produced offline. This was examined in depth to determine whether the second rater deviated from the first rater. Since the FMA uses quite a broad rating scale (0, 1, or 2) and because the second rater was trained by the first rater, only a few deviations in score (< 5) were observed and corrected.

3.7.5 Online Data Collection

The data collection phase required a laptop, running the Data Saver GUI, to be set up at the subject's bedside. The laptop was later replaced by a tablet for ease of testing. The Data Saver GUI was used to form Bluetooth connections with the three data logger

boards. This enabled a stream of data to be transmitted between the data logger boards and the laptop/ tablet.

Prior to subject evaluation, important testing parameters were collected and supplied to the GUI. These include the subject Unique IDentifier (UID), active ADC pins of the data logger board, and side of the body instrumented. Additional information including the data logger board number and the sampling rate was provided automatically by the data logger board following successful Bluetooth connection. An image of the save tab if the GUI is shown in Figure 3.24.

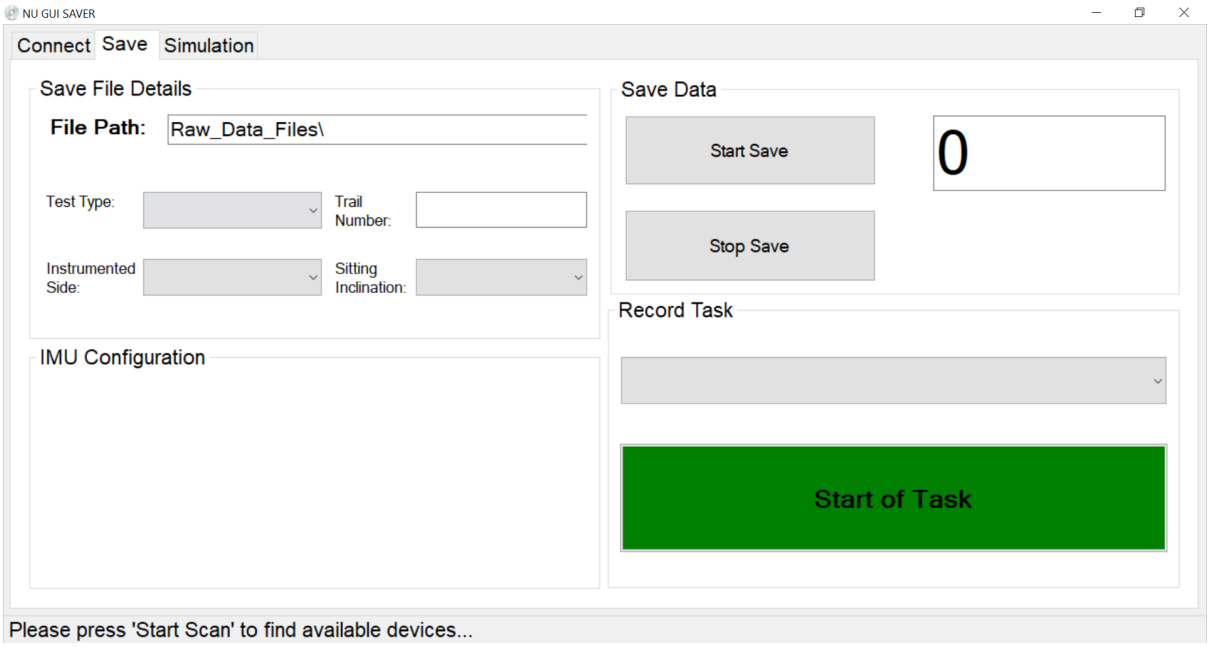


Figure 3.24: Data Saver GUI- Save Tab

Once the subject was prepared for clinical evaluation, the "Start Save" button of the GUI was pressed. Digital markers were manually assigned in the data by a button press during the saving process. These markers served to indicate the start and end of each task and were later used to segment the data. This is an important step since subsequent analysis required the data for each individual task so that the rating could be predicted on a task-by-task basis across all subjects. Once clinical evaluation had been completed, the "Stop Save" button was pressed. All data was saved locally in the Comma-separated Values (CSV) format with a single file generated for each trial. In addition to saving the data, the GUI also exposed functions for displaying plots or a real-time avatar (as discussed in greater depth later in the chapter).

The final set of data consisted of a CSV data file for each of the 64 subjects tested. Each of these files was organised by the task performed using the digital markers. This was an important step since the classification model discussed later was used to classify the labels (clinical scores) assigned to the task using the task data only. This meant that the data corresponding to the same task had to be extracted from each and every subject tested. CSV files were formatted so that all necessary data about the test (anonymised patient number, sampling rates, etc) was accessible and data could easily be processed in subsequent stages.

A summary of all the major components of data collection pipeline is shown in Figure 3.25. Examples of the task segmented data streamed from each axis (and overall) of the accelerometer and from the MMG sensors are shown in Appendix B Figure 9.1 and Figure 9.3 respectively.

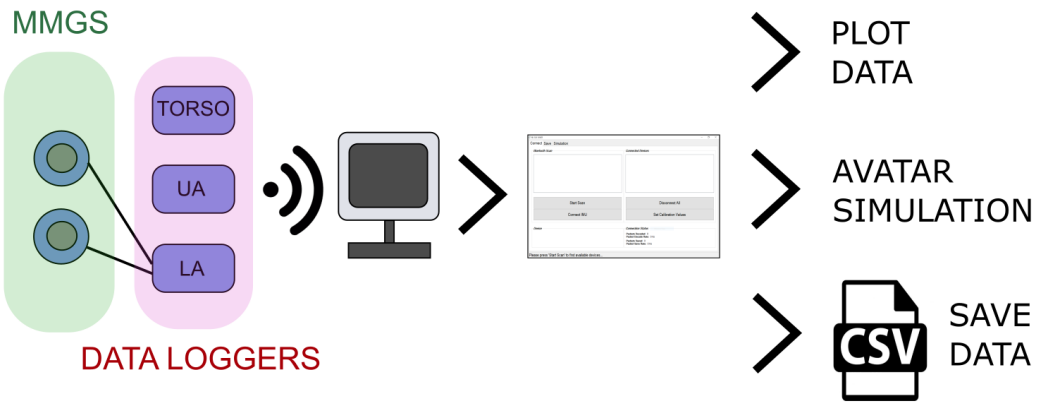


Figure 3.25: Project data pipeline. Data is sampled from the data logger boards and transmitted via Bluetooth to the computer

3.8 Data Processing

All major data processing steps were conducted using the Data Processor GUI. A distinction was made between local and global-scale processing for this study depending on the dataset being operated upon. Local processing defines all operations that were performed on the individual datasets captured from **each** subject trial. This includes operations which were applied to each instrumented limb segment such as filtering and orientation calculations. Global processing defines the operations that were subsequently

performed on the concatenated dataset across **all** subject trials. Concatenation of the dataset was performed to maximise efficiency and minimise the risks of memory errors.

The importing of the CSV file, generated during data collection, was handled by the "Select Data" tab of the Data Processor GUI (see Figure 3.18). In addition to the data CSV files, other files were provided corresponding to the tasks performed, trial specifications (sampling rates, sensor locations), calibration specifications, and the clinical rating data.

3.8.1 Local Data Processing

The main stages included in local processing phase were calibration and standardisation of raw data, quaternion calculation, orientation taring and offset, and data filtering. An image of the local processing tab of the Data Processor GUI is shown in Figure 3.26.

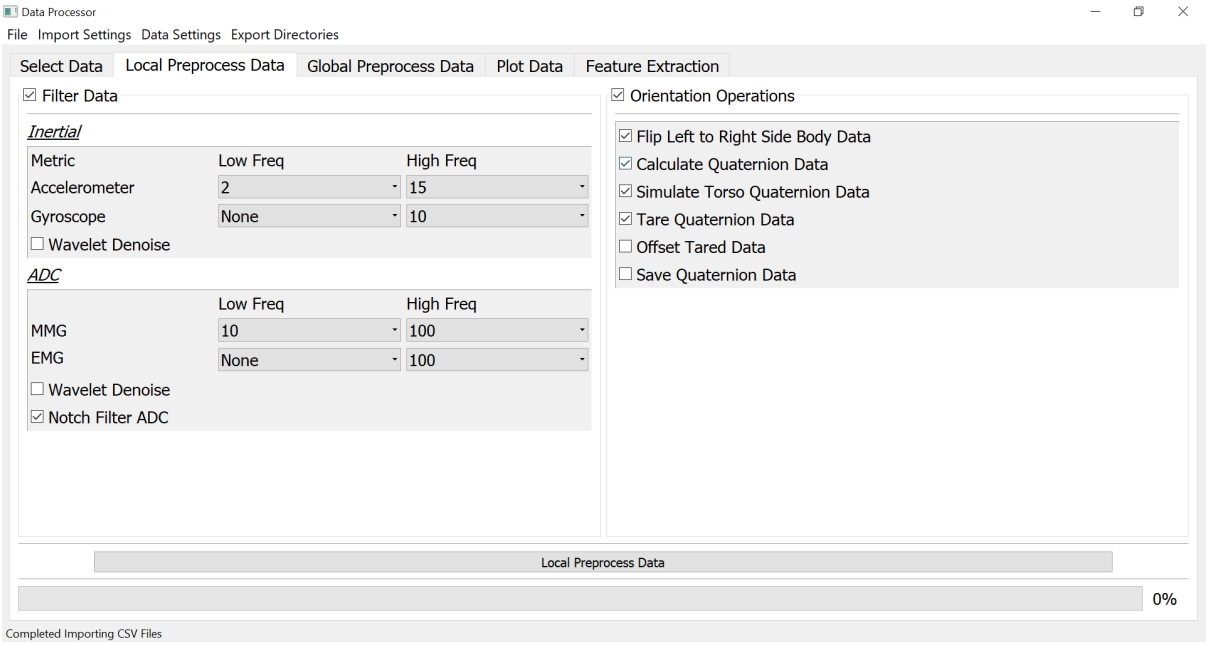


Figure 3.26: Data Processor GUI- Local Process Tab

The first stage of local processing was the calibration and standardisation of the data. Bias values for each axis of the accelerometer and gyroscope for each data logger board were collected by calibration testing prior to any subject testing. Magnetometer calibration values were collected before testing each subject since the appropriate bias values tended to drift rather than remain constant. The calculated calibration values were imported into the GUI and applied to each sensor axis to remove the bias from the

data. The data was subsequently normalised by converting the inertial and ADC data recorded into the following standardised units:

- Accelerometer: Gravity (g)
- Gyroscope: Radians per second (rad/s)
- Magnetometer: Gauss (G)
- Mechanomyogram: Voltage (V)

Quaternion data was calculated using a Python coded version of the Madgwick Gradient Descent Algorithm (see Subsection 2.8.3: Madgwick Gradient Descent for derivation and performance of this algorithm). A modified version was used which required only the accelerometer and gyroscope metrics due to issues with the drift of the magnetometer in this study. Prior to calculating the quaternion values, the sensor axes were rotated to match the expected axes of the Gradient Descent algorithm (y-axis of each sensor metric parallel to gravity in a null rotation orientation). Another step was made to flip certain axes of data recorded on the left-hand side of the subject to simulate recording on the right-hand side. This meant that it would be possible to directly compare recordings taken of subjects regardless of the side of the body which was instrumented.

Orientation taring and offset was a necessary step to apply to the calculated quaternion data. Magnetometer data was not implemented in the Madgwick Gradient Descent Algorithm due to bias issues and magnetic interferences. As a result, there were not the two fixed vectors required by the Gradient Descent algorithm to find a unique orientation, and this resulted in heading drift. This clearly presented an issue for this study since this heading drift would increase over time during the trial and eventual result in the orientation computed for each IMU becoming out of sync. The solution developed for the study was to take advantage of known, fixed start positions with which to position the test subject at the beginning of each task. This meant that the orientation could be tared at the start of each task and as a result there would only be a very small amount of heading drift over the short time course of a single task. There were two main methods of taring each task proposed for this study:

1. For tasks which only involved rotation around a single joint axis or minimal discernible rotation, the orientation was fully tared to simulate the appropriate starting position. This method involved resetting the orientation completely to a known start position for the given task. All of the tasks included within the hand and wrist category of the FMA were tared using this method.

2. For complex tasks involving rotation around multiple joint axes, it was not considered to be reasonable to assume that every subject would be able to get into the correct start position. For instance, for the extensor synergy task, it is necessary for the subject to start with their hand on their ear. This position is clearly not feasible for many subjects. Since drift was only present around the heading axis, a solution was reached by only taring around the heading axis (using swing-twist decomposition). Small manual orientation adjustments were then made to these tasks using custom heading offset angles to ensure that these heading orientations were correct.

The final local processing step was the filtering of the data. All prior operations on the sensor data required it to be unfiltered. For all subsequent operations, the data had to be filtered to remove the noise present in the signal. All data were filtered using a Butterworth Infinite Impulse Response (IIR) filter (SciPy Library). Acceleration was band-pass filtered at 2-15 Hz while the gyroscope was low-pass filtered at 10 Hz. The MMG signal was band-pass filtered at 10-100 Hz. This frequency range was chosen since it removes most of the low-frequency motion artefacts and high-frequency noise [91] while preserving the majority of the useful signal [92].

3.8.2 Global Data Processing

Global processing was administered after raw data had been locally processed and concatenated into large dataframes encompassing all trials conducted. The main stages of global processing were the initial concatenation of trial data, derivation of joint metrics, calculation of data magnitude, and the calculation of ratios between the different metrics. An image of the global processing tab of the Data Processor GUI is shown in Figure 3.27. The first stage of global processing was to concatenate the data collected from each trial into a single large dataframe. This stage was critical to enable subsequent operations to be performed in a time-efficient manner. To enable this concatenation, data recorded from each trial was first resampled to match the lowest sampling rates recorded across all trials. This corresponded to a sampling rate of 100 Hz for the inertial data and 500 Hz for the MMG data.

Joint metrics for the shoulder and elbow were calculated from the quaternion data available for each body segment. Joint angle magnitude was calculated by finding the angle between a vertical vector (v_o) and the same vector rotated by the joint quaternion (q_j).

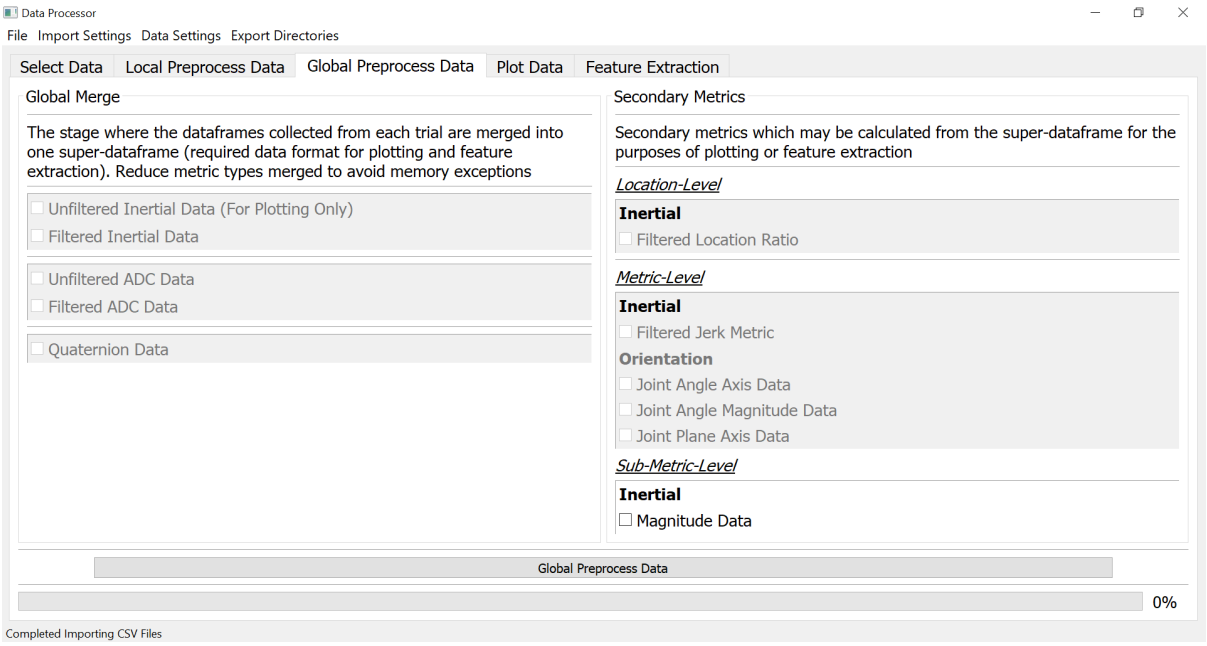


Figure 3.27: Data Processor GUI- Global Process Tab

The minimum angle between these vectors was calculated from the 2-argument arctangent of the norm of the cross product and the dot product as shown in the equation below:

$$v_o = [0, 0, 1]$$

$$v_r = q_j \otimes quat(v_o) \otimes q_j^{-1}$$

$$r_{mag} = arctan2(norm(cross(v_0, v_r)), dot(v_o, v_r)) \tag{3.1}$$

Three additional joint angle metrics were calculated to supplement the joint magnitude. These were the joint flexion, joint abduction, and joint twist. Shoulder flexion and abduction were calculated by rotating a vertical vector by the joint quaternion. The degree of shoulder flexion or abduction could then be determined based on how strongly the vector aligns with a particular plane (the x or y plane in this case). Joint twist angle for both the shoulder and elbow was calculated by the use of swing-twist decomposition. This was used to decompose the twist component of the joint quaternion around a vertical vector. The rotation may then be quantified as a vector rotation by quaternion multiplication of an orthogonal vector by this twist quaternion. The twist angle was then found by finding the minimum angle between the original and rotated vector using the same method as in equation (3.1). The equations for calculating the joint flexion and joint abduction are given in equation (3.2) and equation (3.3) respectively. The twist quaternion (q_{twist})

for a particular joint (shoulder or elbow) was found using the swing-twist decomposition algorithm (equation (2.11)). The equation for the calculation of the twist angle (r_{twist}) is shown in equation (3.4). Graphs of the calculated joint metrics for the shoulder and elbow for the "Shoulder Flexion" task are shown in Appendix B Figure 9.4.

$$r_{flex} = v_r[0] * 90 \tag{3.2}$$

$$r_{abduction} = v_r[1] * 90 \tag{3.3}$$

$$v_{twist} = q_{twist} \otimes quat(v_o) \otimes q_{twist}^{-1}$$

$$r_{twist} = arctan2(norm(cross(v_0, v_{twist})), dot(v_o, v_{twist})) \tag{3.4}$$

The magnitude was calculated for both the accelerometer and gyroscope measures. This provides a measure of the energy over the given metric which is invariant to the direction the task was performed. This is useful since the subject may perform the task incorrectly or use compensatory movements which could result in misleading features when deriving from the axes measures only. This metric may be considered to provide an instantaneous measure of the intensity with which a subject is performing a given action [63]. The magnitude over the three axes was calculated using the Euclidean norm across the axes, as shown below for acceleration:

$$|a| = \sqrt{a_x^2 + a_y^2 + a_z^2} \tag{3.5}$$

The ratio between metrics was the final global measure to be calculated. The ratio was found between the upper arm (UA) and the lower arm (LA) for the magnitude of acceleration and gyroscope as well as the axes of the gyroscope. A simplified version of the ratio equation for the magnitude of acceleration is shown below. A minimum value requirement for both operands was implemented to avoid misleadingly high ratio values:

$$Ratio(|a|_{UA}, |a|_{LA}) = \left\{ \begin{array}{ll} \frac{|a|_{UA}}{|a|_{LA}}, & \text{if } |a|_{UA} * |a|_{LA} \geq threshold \\ 1, & \text{if } |a|_{UA} * |a|_{LA} < threshold \end{array} \right\}$$

3.9 Data Visualisation

Data visualisation was an important part of this project for both online and offline exploration of the data. Keys functions were to assess any errors in data collection and processing, and the identification of useful features. Data visualisation conducted for this study may be categorised as graph visualisation and avatar visualisation.

3.9.1 Graph Visualisation

Graph visualisation was performed both online (using the Data Saver GUI) and offline (using the Data Processor and Feature Explorer GUIs) but was predominantly used offline. The Data Processor GUI was used to plot the raw and processed metrics (including orientation data). Data could be plotted as a whole dataset segmented for each task or as a plot per task. This enabled a high level of detail when checking for discrepancies in task segmentation or sensor errors. A range of plotting options are exposed to enable supplementary metrics to be provided to the plots. These include the addition of a grid or the RMS lines of the data. An image of the plot settings given in the "Plot Data" tab of the Data Processor GUI is shown in Figure 3.28. Graph visualisation was also performed using the Feature Explorer GUI for the purpose of assessing feature correlation and usefulness and discussed later in this chapter. A selection of the plots generated using the Data Processor GUI are shown in Appendix B.

3.9.2 Avatar Visualisation

Avatar visualisation was developed as a function of the Data Saver GUI for both online and offline viewing. Model visualisation was performed using the OpenTK library. This library provides low level C# bindings for OpenGL. Operations for online and offline visualisation were the same with the exception that for offline visualisation pre-recorded quaternion data had to be synchronised and curtailed across each limb segment first. In addition, information had to be provided concerning the location the quaternions were recorded from (torso, upper arm, or lower arm) as well as the side of the body that was instrumented. Starting limb locations were selected so that the avatar was in anatomical position (with the exception that the arm is pronated instead of supinated).

The first quaternion operation was to convert the quaternions, calculated from the inertial data at each limb segment, from the Madgwick to the OpenGL reference frame

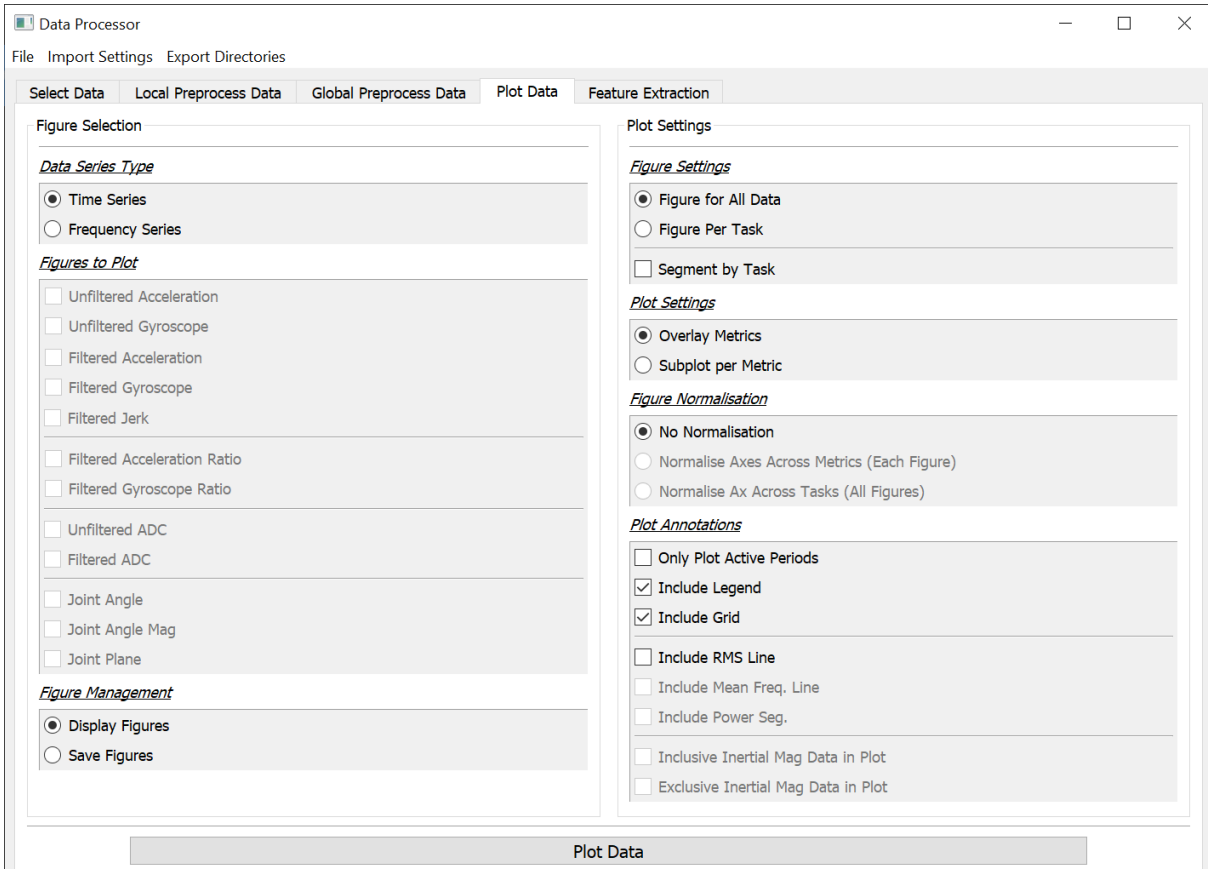


Figure 3.28: Data Processor GUI- Plot Data Tab

(since this is the axis frame required for visualisation). The equation for this axis transformation is outlined below. Superscripts are used to represents the frame of reference and include the Madgwick (M) and openGL (GL) axis frames. The subscripts represent the body segments of the torso (T), upper arm (UA), and lower arm (LA).

$$q_T^{GL} = q^{M \to GL} \otimes q_T^M \otimes (q^{M \to GL})^{-1} \tag{3.6}$$

Following the calculation of the orientation for each body segment, the next step was to find updated body positions (P). An exclusion to this was the torso since this was considered to be fixed in space for ease of visualisation. The position update equations for each of the limb segments are as follows, where the superscript "origin" defines the original position the segments were defined in space.

$$P_T = P_T^{origin} \tag{3.7}$$

$$P_{UA} = P_T + (q_T \otimes (P_{UA}^{origin} - P_T^{origin}) \otimes q_T^{-1}) \tag{3.8}$$

$$P_{LA} = P_{UA} + (q_{UA} \otimes (P_{LA}^{origin} - P_{UA}^{origin}) \otimes q_{UA}^{-1}) \tag{3.9}$$

Once the updated limb segment orientations and positions had been derived, these could be applied to a series of 3D anatomical models to generate an avatar. Avatar models were produced for each of the 64 trials conducted. These visualisations were subject to the taring operations applied to the orientations calculated for each IMU, for instance taring to a fixed start position or just around the heading axis (see Subsection 3.8.1: Local Data Processing). The produced avatars were critical for ensuring that orientation calculations were performed correctly and for ensuring the tasks were tared correctly. Secondly, the avatars were used to ensure consistency of scores between the two clinicians responsible for assigning the FMA-UE ratings (see Subsection 3.7.4: Clinical Assessors and Training). An image of this avatar visualisation taken at the end of the "Flexor Synergy" task is shown in Figure 3.29.

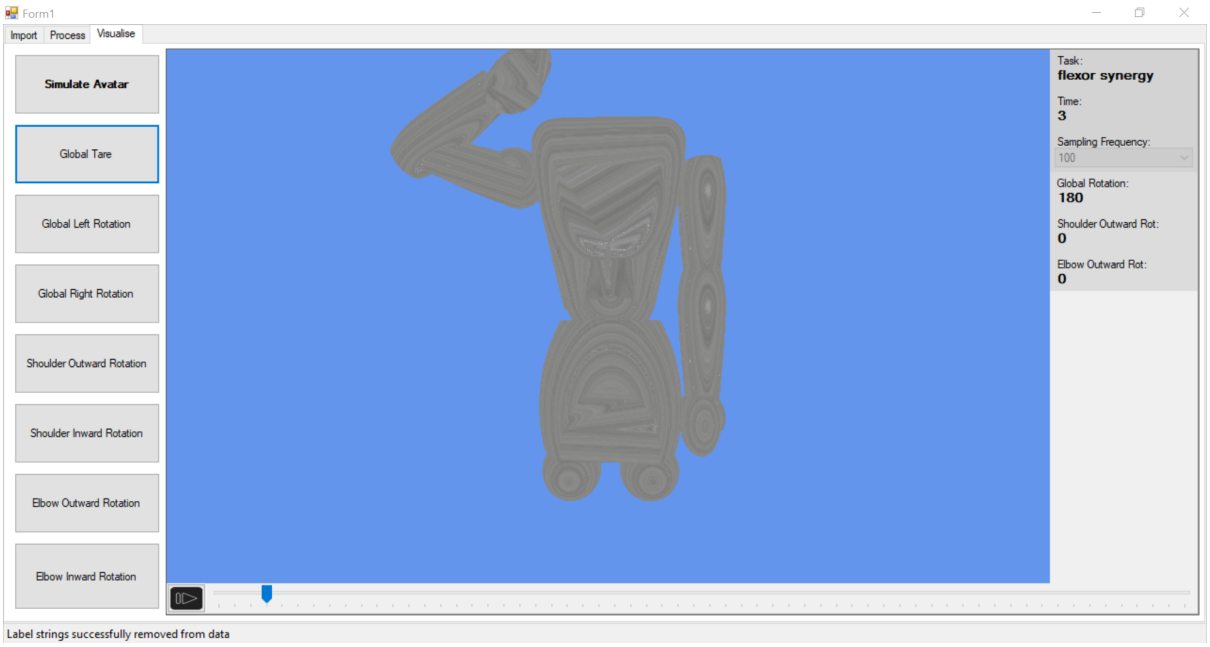


Figure 3.29: Data Saver GUI- Avatar simulation using pre-recorded data of the "Flexor Synergy" task of the FMA-UE

3.10 Feature Extraction

This section outlines the actual features which were extracted from all of the sensor data collected from the wearable sensor system. This includes the actual processing steps, the reasoning behind the choice of features, and a more in-depth examination of the calculation steps.

3.10.1 Protocol

Feature extraction for this study was performed using functions exposed by "Feature Extraction" tab of the Data Processor GUI (see Figure 3.30). Settings available include the complexity of the features to output (may be set to None, Simple, or All). This may be adjusted for the time or frequency-series of the inertial and ADC (MMG) data and the orientation time-series data. An option was also provided for selecting the export format (pickle or parquet). This was included due to data errors incurred when using pickle saving format when the feature sets became prohibitively large. Finally, there were options provided for how the data should be windowed. For the present study, the decision was made to window the data over the entire task periods. There were no tasks present in the FMA-UE task-set which could clearly be further segmented into sub-tasks. In addition, windowing over the entire task simplified the data collection (for the clinical and subject) and data processing steps considerably.

3.10.2 Feature Choice

This subsection details the reasoning behind the choice of the inertial, orientation, and MMG features selected for this study. Features were predominantly selected based on those shown to be useful in past studies, and as such this subsection will refer to the papers referenced in the literature review, in particular see Subsection 3.4.4: Feature Calculation and Subsection 3.4.1: Myographic Capture.

For the inertial data, the first step was to compute a variety of basic (and fast to calculate) features for both the acceleration and gyroscope data. These features were the mean, mean absolution value, minimum, maximum, peak to peak, and standard deviation. It is worth noting a few of the aforementioned features were redundant for certain metrics and therefore removed. This included the minimum value calculated for the magnitude of data (since this value generally registered close to zero) and the mean

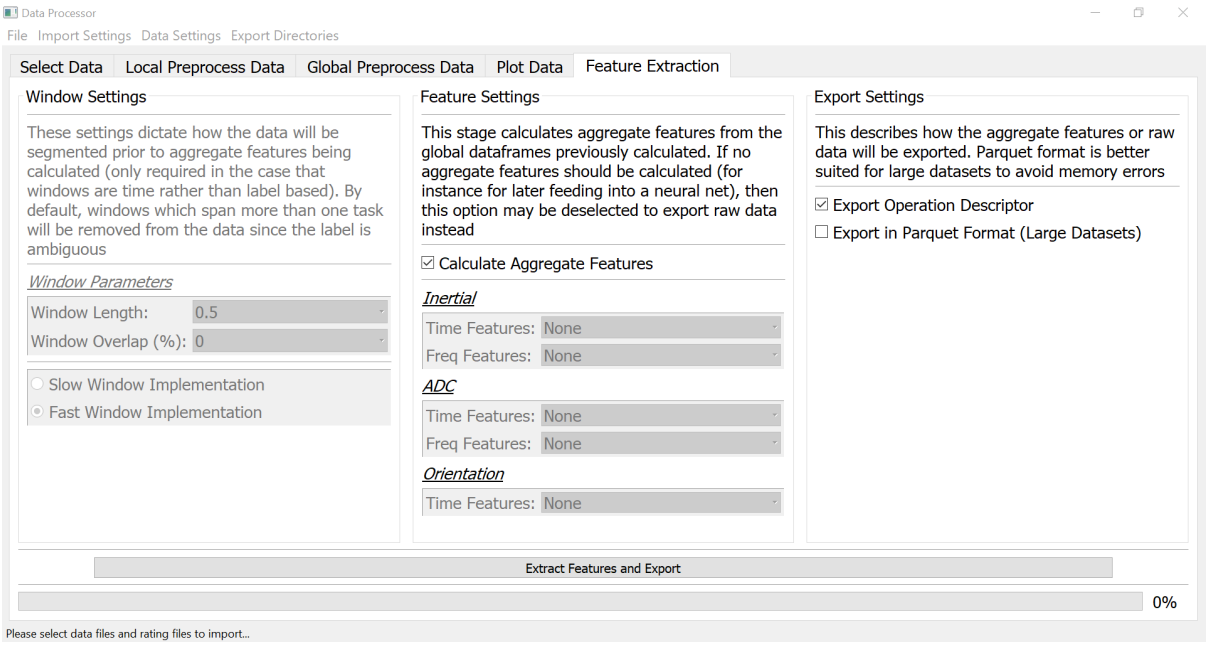


Figure 3.30: Data Processor GUI- Feature Extraction tab

value for accelerometer data since this metric has already been high pass filtered. A series of more sophisticated, higher-order features were computed based on domain knowledge and the features selected in prior studies. The studies by Huang et al [64] and Bento et al [12] were particularly informative since they used IMU configurations similar to the present study.

Orientation features were extracted from the derived quaternion data, as described in Subsection 3.8.1: Local Data Processing. The study by Lee et al [20] focussed on the development of orientation features to quantify tasks of the FMA-UE. This study achieved high classification accuracies by calculating features from joints of the shoulder and elbow, as well as forearm pronation/supination. These features were range of motion, mean, and standard deviation. These features were also employed in the present study, in addition to the minimum and maximum joint angle that the subject could achieve.

Myographic features were calculated from the MMGs placed on the forearm to capture the finger flexors and flexor carpi radialis. There is a limited body of literature available on the use of myographic features to classify motor function, and this is even more limited for the use MMG as opposed to EMG. The clustering index has been found to be statistically significant in distinguishing post-stroke differences in EMG studies but was not employed in this study due to lack of testing for MMG, and the computational difficulties of applying this calculation across irregular time windows. Other features which have shown validity for this application are the RMS and MPF, and both these features

have been implemented in this study. Finally, a number of additional myographic features were selected for this study based on those proposed by Phinyomark et al [93]. This work proposes a large number of time and frequency series features which are considered to be useful for sEMG classification. These features were chosen despite being designed for EMG because no similar work exists for MMG and because the EMG envelope signal may be considered to follow a similar form to that of MMG. One time-series feature proposed by Phinyomark et al which was found to be particularly useful in the present study is the absolute standard deviation of the signal. This suggests that the variance of the MMG signal correlates well with level of impairment of motor function post-stroke.

3.10.3 Feature Computation

Feature extraction was performed on all the base and higher-order metrics calculated over the course of the study. The base measures are the filtered inertial and myographic data. Higher order derived metrics have been computed in the form of magnitude, ratio, and joint orientation. For the sake of brevity, not all features calculated will be covered in this section, but a subsection of the more interesting and useful features will be discussed.

Inertial Time Series Features

The inertial time-series features for the present study were predominantly selected from previous automated studies. A sub-selection of the time series features implemented are as follows:

- **Skew**: A measure of the skewness of a distribution defined as:

$$skew = \frac{m_3}{m_2^{3/2}} \text{ where } m_i = \frac{1}{n}\sum_{i=1}^{n}(x[i] - x)^i \tag{3.10}$$

- **Normalised Median Crossing**: The number of times the signal crosses the median line defined as:

$$S(x_1, x_2) = \left\{ \begin{array}{ll} 1, & \text{if } (x_1 - \tilde{x})(x_2 - \tilde{x}) < 0 \\ 0, & \text{if } (x_1 - \tilde{x})(x_2 - \tilde{x}) \geq 0 \end{array} \right\}$$

$$mzc = \frac{1}{n-1}\sum_{i=1}^{n-1} S(x_i, x_{i+1}) \tag{3.11}$$

- **Trapezoidal Rule**: A measure of approximating the definite integral defined as:

$$tr = \Delta x \left(\sum_{i=1}^{n-1} x_i + \frac{x_n + x_0}{2} \right) \tag{3.12}$$

where Δx = sample period = 0.01

Inertial Frequency Series Features

A frequency-domain signal complete with frequency bins (b) and amplitude (f) was calculated by applying the Fourier transform to the time signal. A sub-section of the frequency series features subsequently calculated were:

- **Dominant Frequency**: The frequency component of the Fourier transformed data which contains the highest energy. Defined as:

$$domfreq = b(f_{max}^2) \tag{3.13}$$

- **Mean Frequency**: An assessment of the centre of the distribution of power across frequencies defined as:

$$meanfreq = \frac{\sum_{i=1}^{n} b(i) * f(i)^2}{\sum_{i=1}^{n} f(i)^2} \tag{3.14}$$

- **Mean Power**: The mean power of the frequency transformed signal, defined as:

$$meanpower = \frac{1}{n}\sum_{i=1}^{n} (f(i)^2) \tag{3.15}$$

- **Power Ratio**: The ratio of the power above and below the mean frequency value, defined as:

$$overthresh(f,b) = \left\{ \begin{array}{ll} f^2, & \text{if } b > meanfreq \\ 0, & \text{if } b \leq meanfreq \end{array} \right\}$$

$$underthresh(f,b) = \left\{ \begin{array}{ll} 0, & \text{if } b \geq meanfreq \\ f^2, & \text{if } b < meanfreq \end{array} \right\}$$

$$powerratio = \frac{\sum_{i=1}^{n} underthresh(f(i),b(i))}{\sum_{i=1}^{n} overthresh(f(i),b(i))} \tag{3.16}$$

- **Power Spectrum Ratio**: Calculates the ratio of the power in a window around the dominant frequency and the power of the rest of the signal. The lower boundary (LB) and upper boundary (UB) define the boundaries of the window. Equation defined as:

$$LB = domfreq - window/2$$

$$UB = domfreq + window/2$$

$$windowpower = \sum_{i=LB}^{UB} f(i)^2$$

$$powerspectrumratio = \frac{windowpower}{\sum_{i=1}^{n} f(n)} \tag{3.17}$$

Orientation Time Series Features

The orientation time-series features for the present study were predominantly selected from previous automated studies, particular the work by Lee et al [20]. A sub-selection of the time series features implemented are as follows:

- **Range of Motion**: Provides an estimate of the joint range of motion in the plane assessed. Defined as:

$$ROM = Max(\theta_{s,e}) - Min(\theta_{s,e}) \tag{3.18}$$

where s, e represent the shoulder and elbow joint respectively

- **Range of Pronation-Supination**: Provides an estimate of the range of pronation-supination of the forearm. Defined as:

$$PS = Max(\theta_{fa}) - Min(\theta_{fa}) \tag{3.19}$$

where fa represents the forearm

Myographic Time Series Features

Myographic feature were largely chosen from the feature set recommended by Phinyomark et al [93] for EMG signal classification. A subset of the time series features selected for the MMG signal were:

- **Log Detector**: Provides an estimate of the muscle contraction force. Defined as:

$$LOG = e^{\frac{1}{n}\sum_{i=1}^{n} log(|x_i|)} \tag{3.20}$$

- **Myopulse Percentage Rate**: A calculation of the overall amount of time the myopulse output exceeds a set threshold. For the present study, a threshold value of one RMS was set as the threshold for sufficient activity. Equation defined as:

$$S(x) = \left\{ \begin{array}{ll} 1, & \text{if } x \geq threshold \\ 0, & \text{otherwise} \end{array} \right\}$$

$$MYOP = \frac{1}{n}\sum_{i=1}^{n}[S(x_i)] \tag{3.21}$$

- **Slope Sign Change Percentage**: The metric provides frequency information by detecting the number of times the signal changes sign. A smaller threshold value of 0.1 RMS was set to remove false positives from signal noise. Equation defined as:

$$S(x) = \left\{ \begin{array}{ll} 1, & \text{if } x \geq threshold \\ 0, & \text{otherwise} \end{array} \right\}$$

$$SSCP = \frac{1}{n-2}\sum_{i=2}^{n-1}[S[(x_i - x_{i-1}) * (x_i - x_{i+1})]] \tag{3.22}$$

Myographic Frequency Series Features

The frequency-domain signal was calculated from the myographic signal using the same method as applied to the inertial signal, with frequency bins (b) and amplitude (f). A feature set was calculated for this domain by once again using the features recommended by Phinyomark et al [93]. A subset of the selected features is as follows:

- **Spectral Moments**: The Spectral Moments (SM) may be used as a method to define the myographic power spectrum. Defined As:

$$SM0 = \sum_{i=1}^{n} f_i \tag{3.23}$$

$$SM1 = \sum_{i=1}^{n} f_i b_i \tag{3.24}$$

$$SM2 = \sum_{i=1}^{n} f_i b_i^2 \tag{3.25}$$

$$SM3 = \sum_{i=1}^{n} f_i b_i^3 \tag{3.26}$$

- **Variance of Central Frequency**: Variance of the central frequency is an important characteristic of the frequency signal and may be defined using the spectral moments. Equation defined as:

$$VCF = \frac{SM2}{SM0} - \left(\frac{SM1}{SM0}\right)^2 \tag{3.27}$$

3.11 Feature Exploration

Feature exploration for this study was largely performed using the specially designed Feature Exploration GUI. This application provided functionality for the visualisation and manual selection of features.

3.11.1 Feature Plotting

A Feature Explorer tab was included in the GUI and enabled the plotting of any combination of two features onto a 2D graph. In addition, the Pearson's Correlation coefficient was displayed in the corner of each plot. This score provided the correlation of each feature with the given clinical rating score (label) as well as the correlation between the two features.

One function of the plots was to identify any features which may have been calculated incorrectly. In addition, plots helped detect any features which were redundant due to low variance, high correlation with another feature, or very low correlation with the labels. An image of the Feature Explorer tab complete with a plot for the maximum shoulder elevation and magnitude of upper arm gyroscope features extracted from the Shoulder Flexion (90-180) task is shown in Figure 3.31.

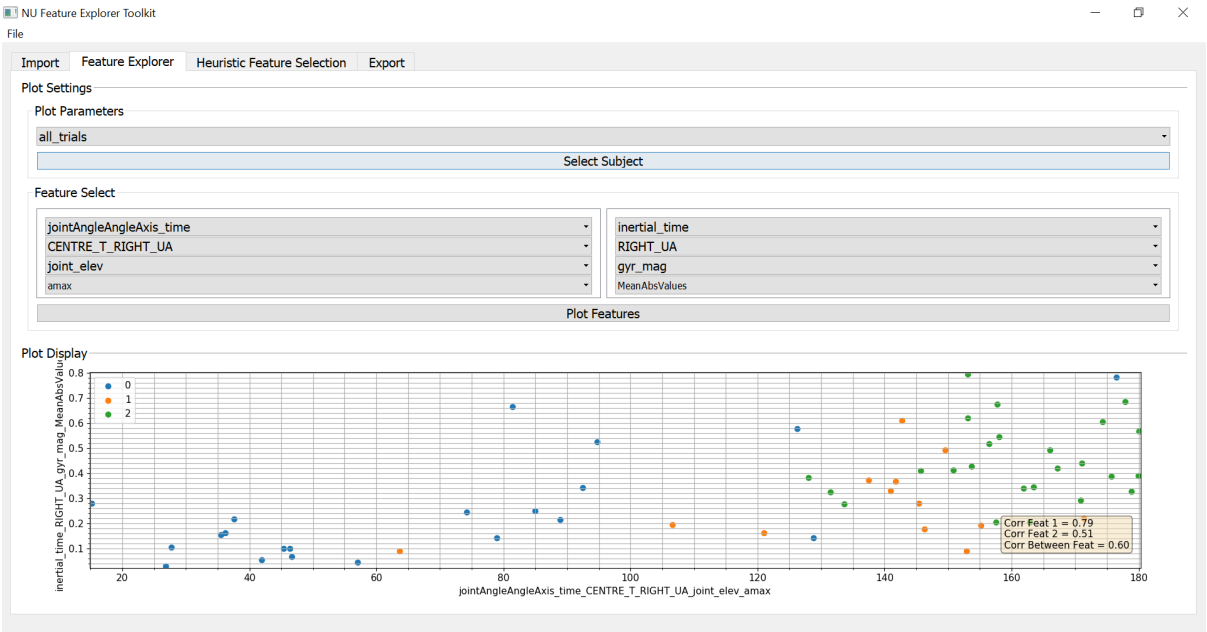


Figure 3.31: Feature Explorer GUI- Feature Explorer tab

3.11.2 Manual Feature Selection

A manual feature selection step was introduced prior to feeding features into the classification pipeline (see Section 3.12: Classification Pipeline). This step was considered necessary since there were several hundreds of extracted features compared to only 64 completed trials. This meant that there was a high likelihood of features that are highly correlated with each other or features that may perform well on the training set by chance. These types of features may not be well detected by automated feature selection and could result in feature redundancy or bias in the resulting classification model. All manual feature selection operations were conducted prior to achieving any classification information or separation of a training and test set to avoid introducing any bias into the study.

Manual feature selection was performed using the Heuristic Feature Selection tab (see Figure 3.32). Options were provided to drop features at any of the four defined hierarchical levels of the feature set. This made it straightforward to drop specific features or all features across a particular metric or location. An option was also provided to drop specific trials from the feature set due to these being recorded incorrectly.

A subset of features was removed based on domain knowledge. This includes the removal of all MMG features for the gross upper arm tasks (since muscle activity of the wrist/hand alone was not found to be a useful feature for the larger movements involving multiple limb segments). For the wrist/hand tasks the features derived from inertial data were removed since these did not capture these movements. Features were also removed due to the results of the Feature Explorer tab plotting of data. This includes the removal of features if they had low usefulness or high correlation with other features (as assessed using the Pearson's Correlation coefficient). Finally, data corresponding to entire task recordings for single subjects were also dropped. This was because some tasks may have had incomplete or bad recordings due to the data being curtailed, sensor error, or subject confusion.

3.12 Classification Pipeline

The Feature Classifier GUI was the final piece of software implemented in the present study. This GUI implemented a classification pipeline that covered all the necessary steps required to prepare the data, classify the feature set, and evaluate classification perfor-

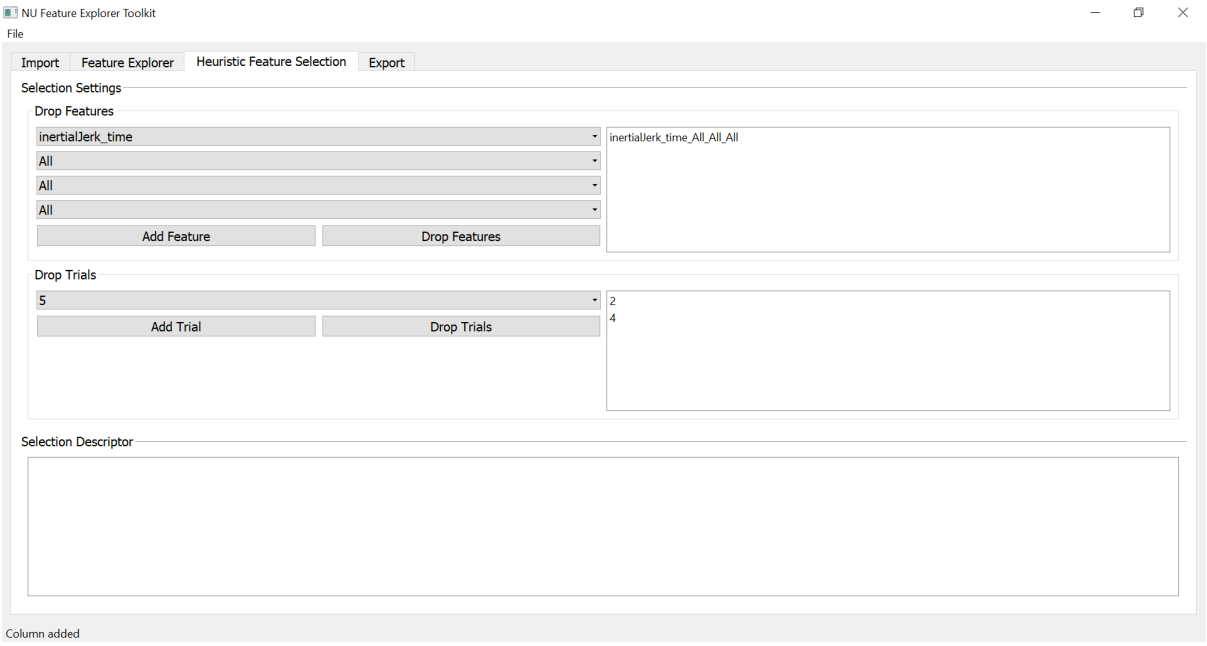


Figure 3.32: Feature Explorer GUI- Heuristic Feature Selection tab

mance. These steps are detailed in the following sections of feature preparation, feature selection, the feature classifier, cross-validation, and evaluation. An image representing the main steps of the classification pipeline is shown in Figure 3.33.

3.12.1 Feature Preparation

The feature preparation stage of the classification pipeline included feature normalisation. For the present study, all features were normalised using standardisation. This is the typical method of feature normalisation and involved subtraction of the mean and division by the standard deviation for each feature.

Resampling of the feature set was a required step due to the biased class ratings for some of the tasks performed. Although every effort was made during this project to test subjects with a wide range of motor function capabilities as possible, imbalances in the difficulties of certain tasks meant that the scores ended up being skewed. To compensate for the imbalance in the dataset, oversampling was used so that each of the classes would be represented equally in the dataset.

3.12.2 Automated Feature Selection

An automated feature selection step was incorporated into the classification pipeline to avoid over-fitting the model to the training set. There were two primary considerations

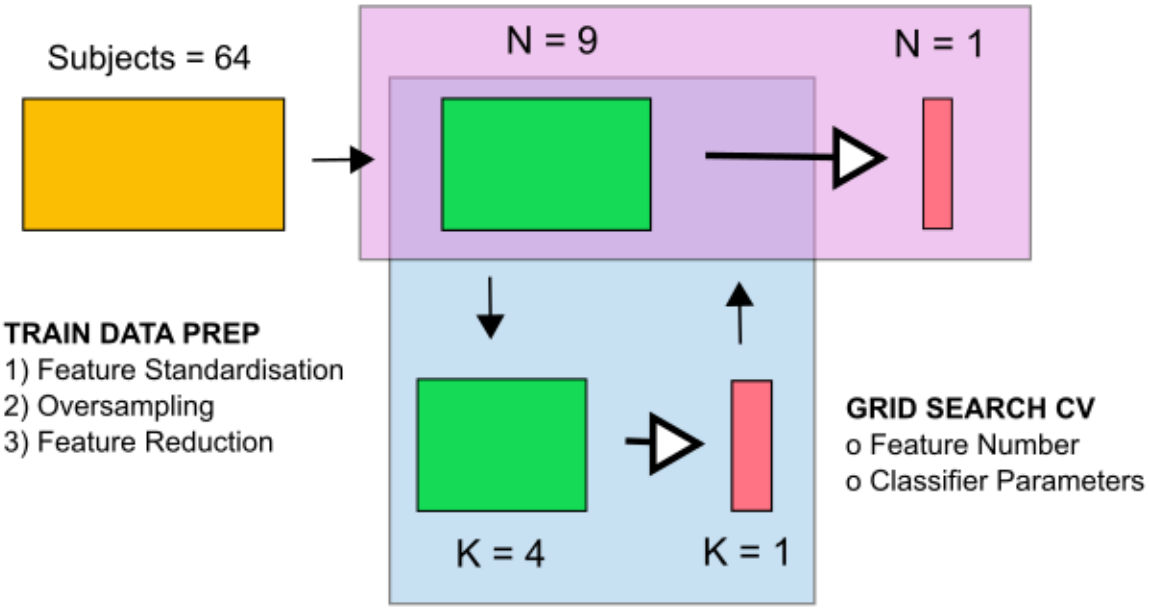


Figure 3.33: Representation of the classification pipeline. Two stages of cross validation shown. External cross-validation (N=10) is shown with a pink background and used for accuracy testing. Grid search cross-validation (K=5) is shown with a blue background and used for optimisation of the hyper-parameters

when optimising the feature selection stage. Firstly, the selection of a feature selection algorithm. For this study, PCA was chosen as the method of dimensionality reduction since it outperformed other selection methods on a with-held test set. Secondly, a choice had to be made on how many features should be preserved. The optimal number of preserved features following PCA was difficult to estimate since this number likely differed for each task to be classified. A solution was implemented by embedding the preserved feature number into the classification analysis as a hyper-parameter. This meant that the optimal preserved feature number from a given subset could always be found. The set of values corresponding to the number of preserved features which could be selected was: [3, 5, 10, 15].

3.12.3 Feature Classifier

The final stage of the classification pipeline was the training, testing, and evaluation of the classifier. For the analysis, the LightGBM classifier was chosen due to its high accuracy for low sample datasets, robustness against over-fitting, and quick training time. Data were split into training and test sets as described in the next subsection. Classifier performance was evaluated by the accuracy as well as the F1-score for each of the classes.

The selection of the classification stages of the pipeline (in addition to feature selection) was incorporated into the Hyper-parameter Classification tab of the Feature Classifier (see Figure 3.34).

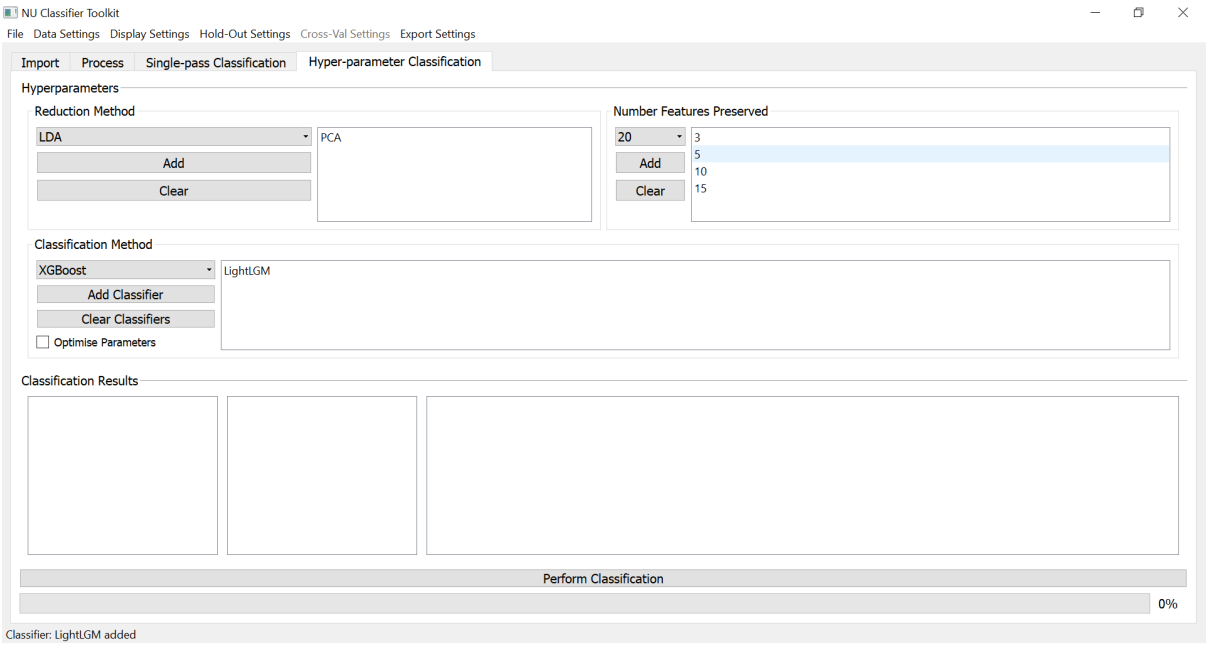


Figure 3.34: Feature Classifier GUI- Hyper-parameter Classification tab

3.12.4 Cross-Validation

For the classification analysis, 10-iterations of 10-fold cross-validation was performed. This method of validation was chosen since it provides a reliable measure of accuracy which does not over-fit the data (like n-1 cross-validation) or lack the necessary training data (such as 2-fold or 5-fold). The folds of cross-validation were stratified to preserve an equal class distribution in each fold. 10-iterations of cross-validation were performed to find more representative classification results. Single iteration analyses are prone to achieving misleading results due to randomness in the way the data may be split depending on the seed of the analysis.

Hyper-parameter optimisation was integrated into each training fold during cross-validation to find the optimal number of preserved features and LightGBM parameters for classification. This was achieved by applying a secondary grid-search cross-validation (5-fold) across each of the training sets (see Figure 3.33 for illustration). This method computes a score for each of the hyper-parameters tested and then uses that parameter

to train and test on the cross-validation train and test sets respectively.

3.12.5 Evaluation

The first goal of the study was to determine the performance of the developed system when classifying the standard ratings as assigned by a clinician for each task. This is the predominant method as performed in the literature and provides predicted scores which are well known and validated. This prediction method was implemented by setting up a classification pipeline with labels provided by the three scores that make up task ratings for the FMA-UE clinical scale (0, 1, or 2). For three of the tasks of the FMA-UE (flexor synergy, extensor synergy, coordination-speed) the combined score of the subcomponents exceeded these ratings. To ensure it was possible to classify these tasks, scores were downsampled to the (0, 1, or 2) rating. Unfortunately, several of the hand/wrist tasks contained very imbalanced datasets due to the typically high ratings achieved for these tasks. Each of these tasks contained > 32 subjects scoring a rating of 2 as opposed to < 9 subjects scoring a rating of 0 or 1. This meant that over 50% of the class labels belonged to the upper class and less than 8% of class labels belong to one of the other two classes. From this distribution it is clear that the classifier would fail to fully learn these poorly represented classes and would likely have a strong bias towards the upper label class. For this reason, the tasks of cylinder grasp, hook grasp, pincer grasp, and spherical grasp were removed from the classification analysis to avoid giving misleading results.

A secondary classification goal was to determine how well the model could distinguish between healthy, and anything from mild to major impairment. This method is less widely reported in the literature but still holds validity as a lower resolution score of motor function for applications including home-based assessment. In addition, it was of interest in the present study how significantly the classification performance could be improved for the two rather than three-class problem. Prediction was performed by combining the two lower scores (0, 1) into a single label representing impaired function $(0, 1) \Rightarrow 0$ and the maximum clinical score (2) being used to assign full function $2 \Rightarrow 1$. This reduced label set (impaired (0) or healthy (1)) was then used to evaluate the two-class classification performance.

The final evaluation step was to assess the statistical significance of the classification results to prove that they were not produced by chance. This was performed for the three class classification model using a repeated permutation method. This method operates by

assuming a condition of independence between the features and classifier. The labels are then repeatedly randomised, and the performance of the classifier trained and tested with the random labels using cross validation. A statistical test is then performed to determine whether a relationship between the features does actually exist and that the classifier predicts the labels better than by pure chance. This would be indicated by a low p-value which would enable rejection of the null hypothesis (that the features and labels are totally independent). It was not possible to implement permutation testing into the classification pipeline developed for this study because each model involved several hyper-parameter optimisation steps and therefore did not have fixed model parameters. A solution is proposed by performing permutation testing post-classification using the most commonly selected hyper-parameters. As used in the classification pipeline, PCA was used for dimensionality reduction and the top five transformed features were preserved. The LightGBM classifier was once again implemented but the default hyperparameters were fixed in this case. The scikit-learn library "permutation test score" was used to perform the statistical testing. This library uses the implementation for statistical permutation testing of classification models as proposed by Ojala et al [94].

3.13 Clinical Feature Set

The first stage of this study was to use the novel instrumented system to develop a classification model capable of distinguishing the rating and impairment level for each task of the FMA-UE. This method is a valid way of approximately a highly validated rating score which does not suffer from subjectivity or require a specialist clinician. The disadvantage of this method is that an approximation of a clinical rating limits the full potential of the accuracy and resolution offered by the instrumented system. To mitigate this limitation, the decision was made for this study to supplement the classified score with some unique features. These features could be used to provide useful feedback to clinician and subject alike and would also be expected to provide a more fine-grained measure of motor function than possible with the clinical rating scales. The combination of these outputs would combine the advantages of reliability and reputation provided by the prediction of the FMA-UE clinical score with the resolution and wealth of data provided by unique features derived from the sensor system.

The criteria for the features included in the clinical feature set were as follows:

1. **Transparency**: It is important the features be as transparent as possible to provide useful feedback to both the clinician and subject alike. This means that the features should correspond to readily understandable aspects of motor function rather than overly esoteric or complex features whose relevance may not be fully understood

2. **Usefulness**: A critical component of the feature is that is must be proven to be a valid and useful measure of motor function. This will be assessed in the present study by correlation with the FMA-UE, since this score acts as the gold-standard measure of measuring upper extremity motor function

3. **Generability**: The feature selected should provide general information about the subject's motor dysfunction and not be task-specific, i.e. not only provide information about how well the subject can perform a given task

3.13.1 Orientation Features

A series of orientation-based measures were selected as clinical features since they may be readily understood and provide clinically sensitive parameters of motor function. These features were based on the maximum or range of joint angles that the subject could achieve. The list of features selected are as follows:

1. Shoulder Rotation: A measure of the subject's ability to rotate their shoulder internally and externally. Measured as the range of shoulder joint rotation exhibited during the "Pronation-supination" task

2. Shoulder Abduction: A measure of the subject's ability to achieve full shoulder abduction. Measured as the maximum shoulder abduction angle recorded during the "Shoulder Abduction" task

3. Shoulder Flexion: A measure of the maximum shoulder flexion the subject can achieve. Measured as the maximum shoulder flexion angle recording during the "Shoulder Flexion (90-180 degrees)" task

4. Elbow Pronation-Supination: A measure of the subject's ability to pronate and supinate their elbow. Measured as the range of elbow joint pronation-supination exhibited during the "Pronation-supination (elbow at 0 degrees)" task

5. Elbow Flexion: A measure of the overall elbow flexion achieved by the subject. Measured as the range of elbow flexion recorded during the "Extensor Synergy" task

3.13.2 Myographic Features

A series of myographic features were chosen which characterised the wrist motor function of the subject. These features were selected since they give a good characterisation of hand/wrist motor function which is otherwise difficult to quantify manually or using conventional wearable sensors. The myographic-based features chosen to supplement the classification score are as follows:

1. Wrist Flexion-extension: A measure of how well the subject can perform wrist flexion and extension. Measured as the difference in absolute standard deviation of the MMG signal during the wrist "Repeated Dorsi/ Volar Flexion" task

2. Wrist Circumduction: A measure of how well the subject can perform wrist circumduction. Measured as the difference in absolute standard deviation of the MMG signal during the wrist "Circumduction" task

3.14 Results

In this section, the results of the automated system developed for this study are displayed. The results are categorised by the two output measures provided by the system: the classified clinical scores and the clinical features. The classified scores have been assessed based on the performance of the classifier. The clinical features have been evaluated based on their usefulness as supplementary metrics to the classified score. Usefulness is determined for the present study by how well the features correlate to the FMA-UE score, since this score forms one of the gold standard measures of motor function.

3.14.1 Classification Performance

Classification performance for this study was assessed based on the classification accuracy and the F1-score achieved for each label. This combination of metrics provides a better picture of classifier performance (particularly classification bias) than accuracy score alone. In particular, the F1-score provides a more in-depth examination of misclassification for each individual label.

Two separate classification models were developed for this study. The first of these was designed using the labels provided by the FMA-UE clinical ratings to output a direct prediction of clinical score. A second classification model was developed which combined the lower two scores of the rating scale to predict healthy or impaired motor function. The performance of both classification models was evaluated using 10 iterations of 10-fold stratified cross-validation. Tasks for these models were separated into gross upper extremity tasks and those which involve the hand/wrist only.

Three-Class Classification

Presented here are the three-class classification performance results for prediction of the ratings assigned to each task of the FMA-UE clinical rating scale (0, 1, or 2). The results for the gross motor tasks and the hand/wrists tasks are shown in Table 3.1 and Table 3.2 respectively.

Two-Class Classification

Presented here are the two-class classification performance results for prediction of impaired or healthy motor function (0 or 1). The results for the gross motor tasks and the

hand/wrists tasks are shown in Table 3.3 and Table 3.4 respectively.

3.14.2 Clinical Features

The clinical features selected to complement the classification score had to be assessed for their usefulness as metrics of motor function. This was performed for the present study by determining the correlation between the features and the local (task) and global (overall) scores of the FMA-UE clinical scale. Since the FMA-UE is the current gold-standard measure of motor function, a good correlation would suggest that the features are clinically valid measures of motor function. However, the FMA-UE also has limitations including low resolution and subjectivity which means that a very strong correlation with this score is unrealistic and potentially undesirable. The correlation results are displayed in Table 3.5. Individual scatter plots for each of the selected features against the global FMA-UE score are shown in Figures 3.35 to 3.41.

Table 3.1: Classification performance results for the novel system as performed on the gross motor tasks of the FMA-UE. Labels classified to are the three labels assigned to each task of the FMA-UE

Task	F1-score (0)	F1-score (1)	F1-score (2)	Accuracy *(p-value)*
Flexor synergy	0.8	0.17	0.86	0.75 *(< 0.01)*
Extensor synergy	0.61	0.26	0.9	0.77 *(< 0.01)*
Hand to lumbar spine	0.64	0.55	0.83	0.72 *(< 0.01)*
Shoulder flexion (0-90)	0.74	0.65	0.9	0.81 *(< 0.01)*
Shoulder abduction	0.62	0.68	0.91	0.81 *(< 0.01)*
Shoulder flexion (90-180)	0.82	0.24	0.8	0.71 *(< 0.01)*
Coordination-speed	0.71	0.54	0.82	0.71 *(< 0.01)*
Mean	**0.71**	**0.44**	**0.86**	**0.75**

Table 3.2: Classification performance results for the novel system as performed on the hand/wrist tasks of the FMA-UE. Labels classified to are the three labels assigned to each task of the FMA-UE

Task	F1-score (0)	F1-score (1)	F1-score (2)	Accuracy *(p-value)*
Pronation-supination (elb. 90)	0.77	0.5	0.86	0.78 *(< 0.01)*
Pronation-supination (elb. 0)	0.51	0.24	0.85	0.69 *(< 0.01)*
Stab. at 15 dorsiflex. (elb. 90)	0.35	0.39	0.75	0.62 *(< 0.01)*
Rep. dorsi-volar flex. (elb. 90)	0.49	0.27	0.82	0.70 *(> 0.1)*
Stab. at 15 dorsiflex. (elb. 0)	0.46	0.41	0.77	0.65 *(> 0.1)*
Rep. dorsi-volar flex. (elb. 0)	0.32	0.18	0.75	0.55 *(< 0.05)*
Circumduction	0.6	0.42	0.69	0.58 *(< 0.01)*
Thumb adduction	0.5	0.32	0.79	0.67 *(< 0.05)*
Mean	**0.5**	**0.34**	**0.79**	**0.66**

Table 3.3: Classification performance results for the novel system as performed on the gross motor tasks of the FMA-UE. Labels classified to are subject performance for each task of the FMA-UE (impaired or healthy)

Task	F1-score (impaired)	F1-score (healthy)	Accuracy
Flexor synergy	0.72	0.86	0.81
Extensor synergy	0.78	0.91	0.87
Hand to lumbar spine	0.78	0.83	0.8
Shoulder flexion (0-90)	0.86	0.9	0.89
Shoulder abduction	0.88	0.91	0.9
Shoulder flexion (90-180)	0.86	0.82	0.84
Coordination-speed	0.82	0.82	0.82
Mean	**0.81**	**0.86**	**0.85**

Table 3.4: Classification performance results for the novel system as performed on the hand/wrist tasks of the FMA-UE. Labels classified to are subject performance for each task of the FMA-UE (impaired or healthy)

Task	F1-score (impaired)	F1-score (healthy)	Accuracy
Pronation-supination (elb. 90)	0.8	0.83	0.78
Pronation-supination (elb. 0)	0.75	0.85	0.81
Stab. at 15 dorsiflexion (elb. 90)	0.48	0.75	0.66
Rep. dorsi-volar flexion (elb. 90)	0.57	0.83	0.76
Stab. at 15 dorsiflexion (elb. 0)	0.51	0.75	0.67
Rep. dorsi-volar flexion (elb. 0)	0.65	0.75	0.71
Circumduction	0.81	0.75	0.78
Thumb adduction	0.61	0.79	0.73
Mean	**0.65**	**0.79**	**0.74**

Table 3.5: Correlation results for the selected features from the novel system. Correlations are made between the features and the local (task) and global (overall) FMA-UE scores

Feature	Corr. with local score	Corr. with global score
Shoulder ext-int rotation (range)	0.47	0.56
Shoulder abduction angle (max)	0.66	0.64
Shoulder flexion angle (max)	0.85	0.78
Elbow pronation-sup (range)	0.77	0.66
Elbow flexion (range)	0.78	0.73
Wrist flex-ext muscle activity	0.52	0.56
Wrist circumduction muscle activity	0.68	0.61
Mean	**0.68**	**0.65**

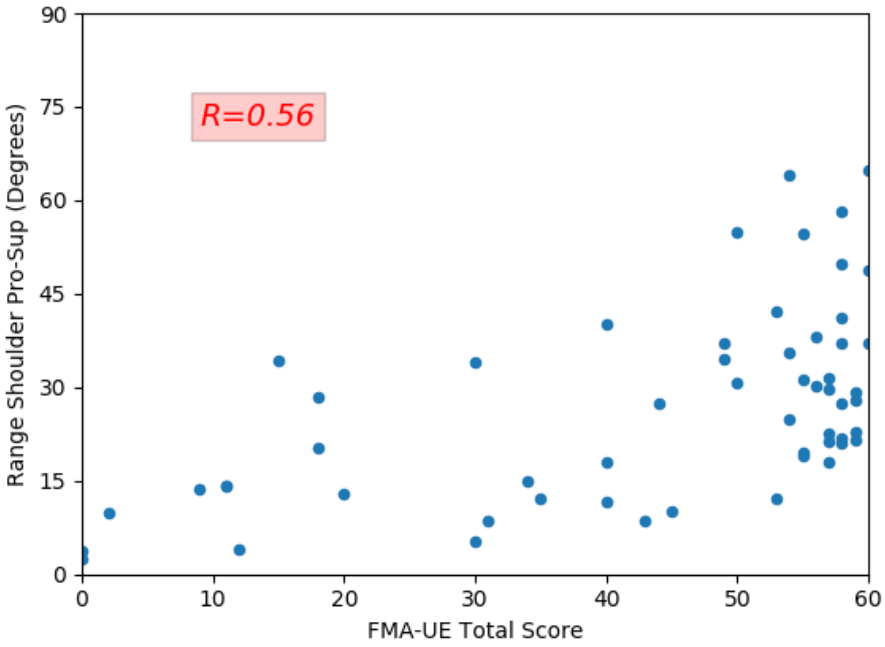


Figure 3.35: Scatter plot of the shoulder ext-int rotation feature versus overall FMA-UE score. Supplemented with Pearson Correlation Coefficient

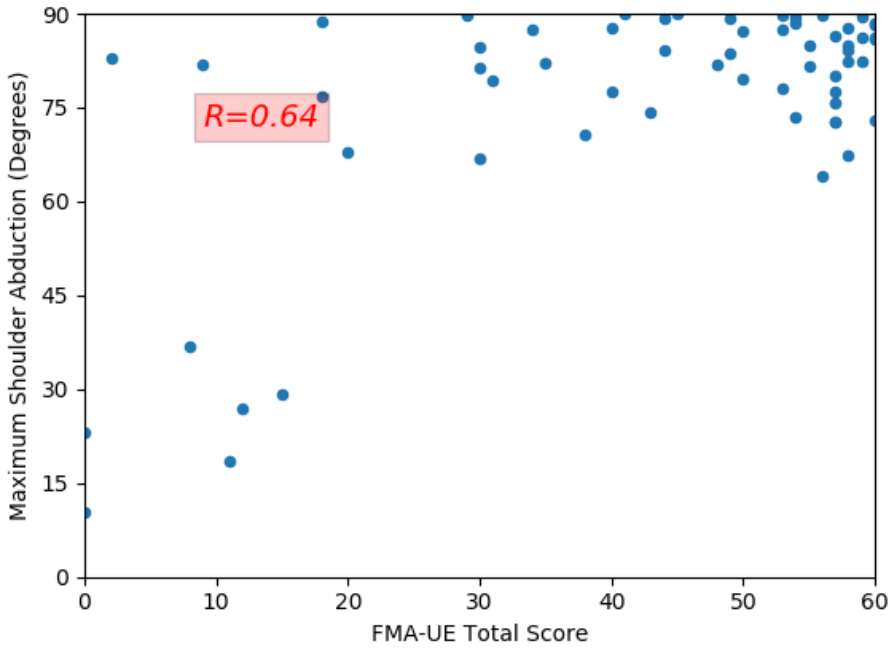


Figure 3.36: Scatter plot of the shoulder abduction feature versus overall FMA-UE score. Supplemented with Pearson Correlation Coefficient

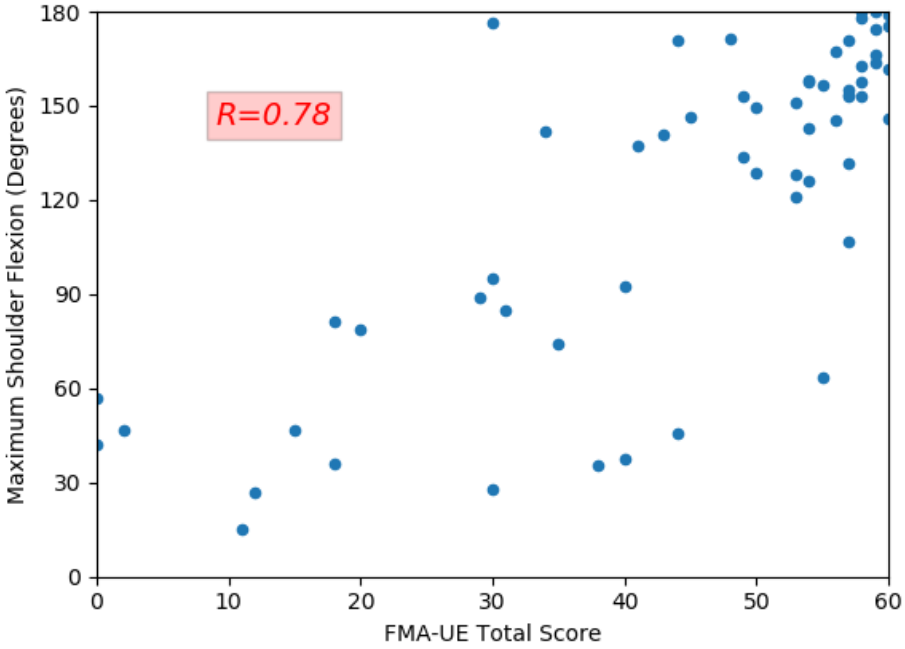


Figure 3.37: Scatter plot of the shoulder flexion feature versus overall FMA-UE score. Supplemented with Pearson Correlation Coefficient

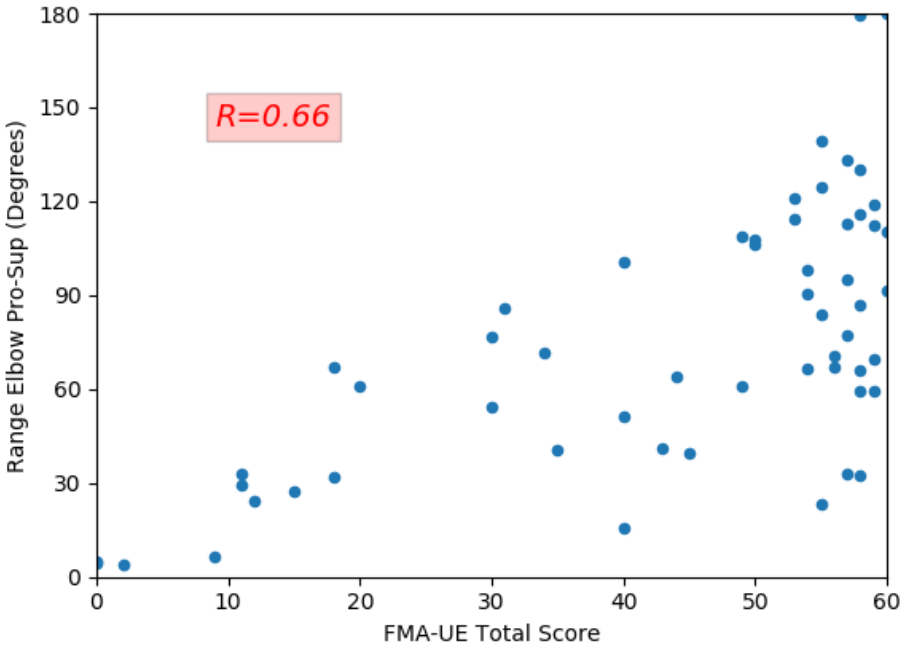


Figure 3.38: Scatter plot of the elbow pronation-supination feature versus overall FMA-UE score. Supplemented with Pearson Correlation Coefficient

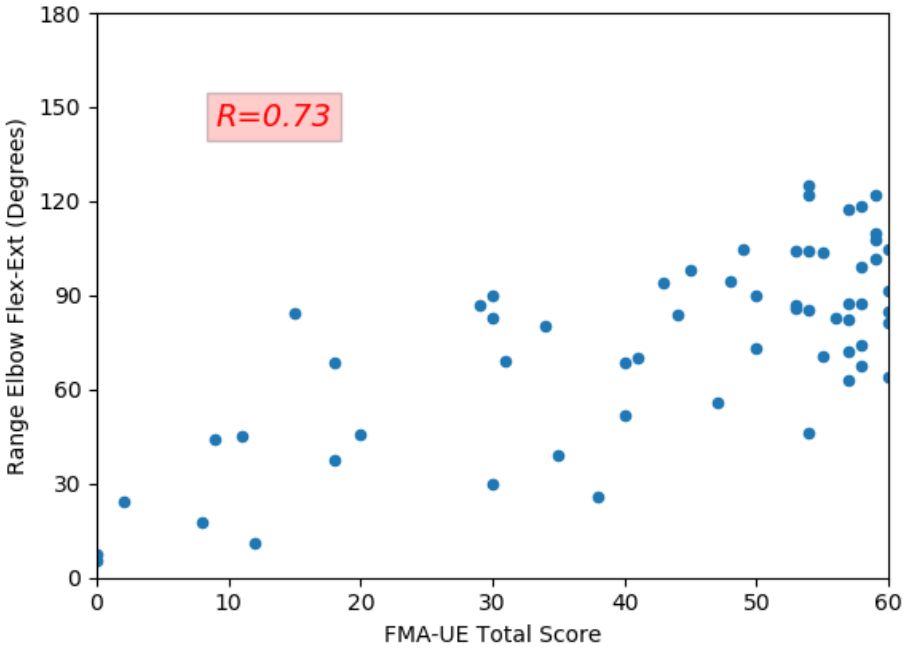


Figure 3.39: Scatter plot of the elbow flexion-extension feature versus overall FMA-UE score. Supplemented with Pearson Correlation Coefficient

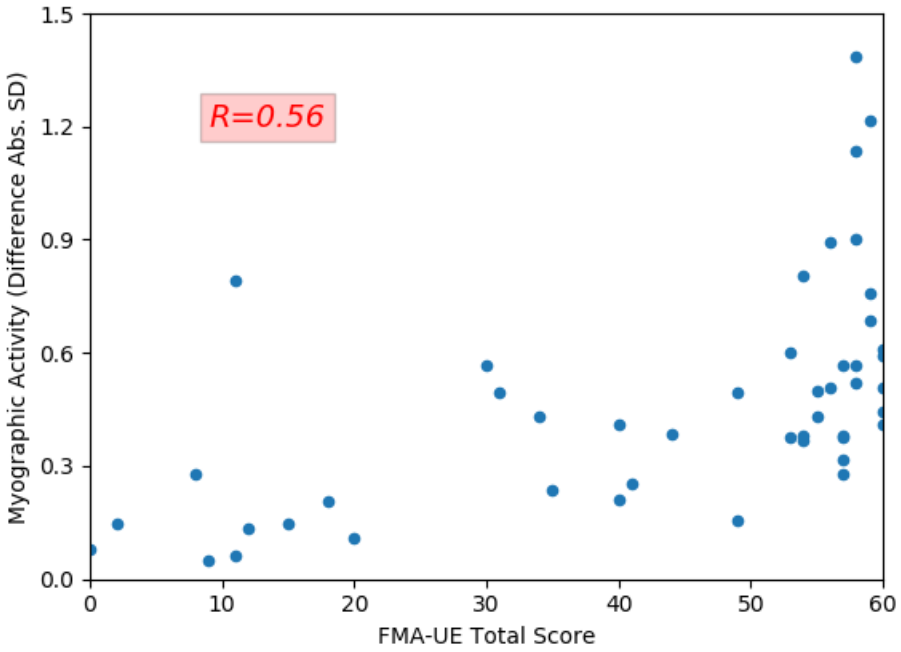


Figure 3.40: Scatter plot of the wrist dorsi-volar flexion feature versus overall FMA-UE score. Supplemented with Pearson Correlation Coefficient

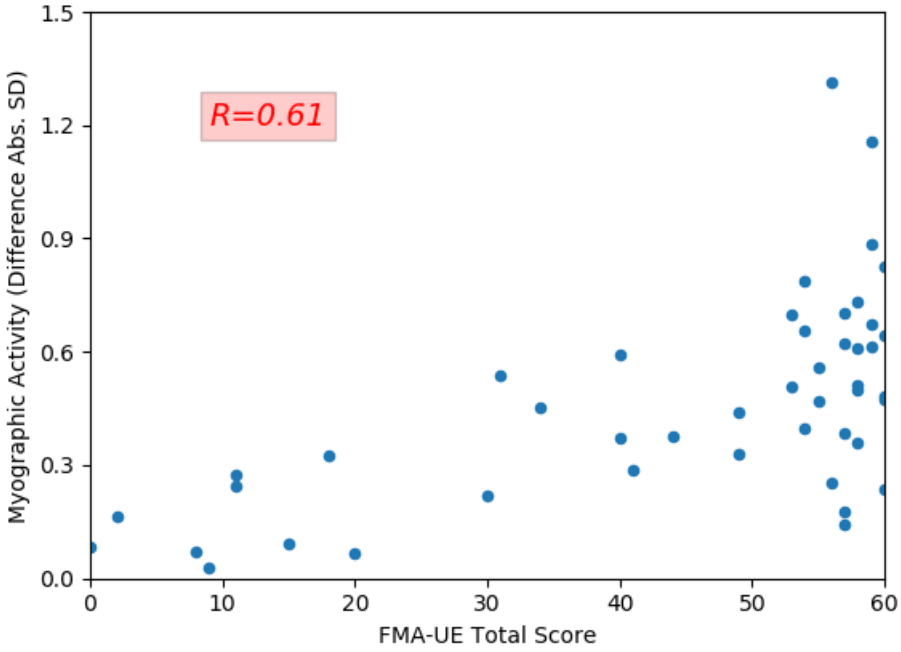


Figure 3.41: Scatter plot of the wrist circumduction feature versus overall FMA-UE score. Supplemented with Pearson Correlation Coefficient

3.15 Discussion

For this section, the results of the automated system developed in the present study will be discussed with respect to meeting the study goals and comparisons with prior automated systems developed in the wider literature. Results are categorised by the two main outputs of the sensor system: the classified rating scores, and the novel clinical features.

3.15.1 Classification Performance

A discussion of the classification results is covered in this section for both the three- and two-class classification models. This is structured as a brief outline of the major findings followed by a comparison of these results with the wider literature. Finally, a more in-depth assessment of the classification findings is discussed with reference to how the parameters of this study may be improved upon.

Three-Class Classification

These classification results were calculated based on the three-labels of the FMA-UE assigned to each task. Results are shown for the gross tasks in Table 3.1 and hand/wrist tasks in Table 3.2. The gross tasks were classified using kinematic and orientation features whereas the hand/wrist tasks were classified using myographic features only.

The mean classification accuracy for the gross motor tasks was 75%. In addition, there was a narrow range of results with a minimum value of 71% and a maximum of 81%. For the hand/wrist tasks, a lower overall classification accuracy of 66% was achieved with a range of 55% to 78%. The F1-score for both sets of tasks indicates that the mid-score (score of 1) was the most poorly classified score and as such contributed most to the loss of accuracy.

There are a limited number of high quality comparable prior automated studies of upper-extremity motor function to compare with the results of the present study. Previous studies were excluded for comparison with the present study based on the following criteria:

- Pilot studies were excluded based on the limited results achievable from a single subject

- Studies were excluded if they performed correlation or calculation of a bilateral metric rather than classification of rating scales due to lack of practical comparison with the present study
- One study [83] was excluded for using healthy subjects to simulate different levels of impairment due to lack of real world application
- Several studies were excluded because they used a generic classification algorithm which was trained with too small a sample size ($\leq$ 15 subjects) to reasonably be expected to generalise to new data

Following this exclusion criteria, there were three studies [19, 20, 67] remaining which met the minimum requirements for comparison with the present study. Despite the suitability of these studies for comparison, there are still several limitations of these studies as compared to the present study which are worth clarifying. These limitations include only classifying a subset of rating scale tasks, an exclusion criterion which excludes bed-bound subjects, and classification results which are generated only over a relatively small sample size.

Kim et al [19] developed and trialled the application of an automated system of the FMA-UE in a large (41 hemiplegic stroke subjects) clinical study. The sensor system utilised was composed of a single Kinect depth-sensing camera. One limitation of the study was that it selected only 13 items from the total tasks involved in the FMA-UE to classify. This is presumably because some of the tasks would be poorly classified by the Kinect sensor such as those involving twist rotation around the bone axis (such as the pronation-supination task) or those involving fine motor function of the wrist or hands. This is unlike the present study which utilised a system capable of classifying all tasks. Another limitation of the work by Kim et al is that only the prediction accuracy is given which means that the classifier may have been biased for certain scores. This would have resulted in a misleadingly high classification accuracy if the dataset were imbalanced. Prediction accuracies achieved ranged from 65% to 87% (see Figure 3.42) although a mean classification accuracy is not provided. These results are comparable to the present study which ranged between 71% and 81% for the gross motor tasks, which cover a similar task set.

The study by Lee et al [20] improved the sensor system set out by Kim et al by the inclusion of force sensing resistor in addition to the Kinect sensor. This enabled the hand grasp tasks to also be quantified. Unlike the present study, a limitation of the work

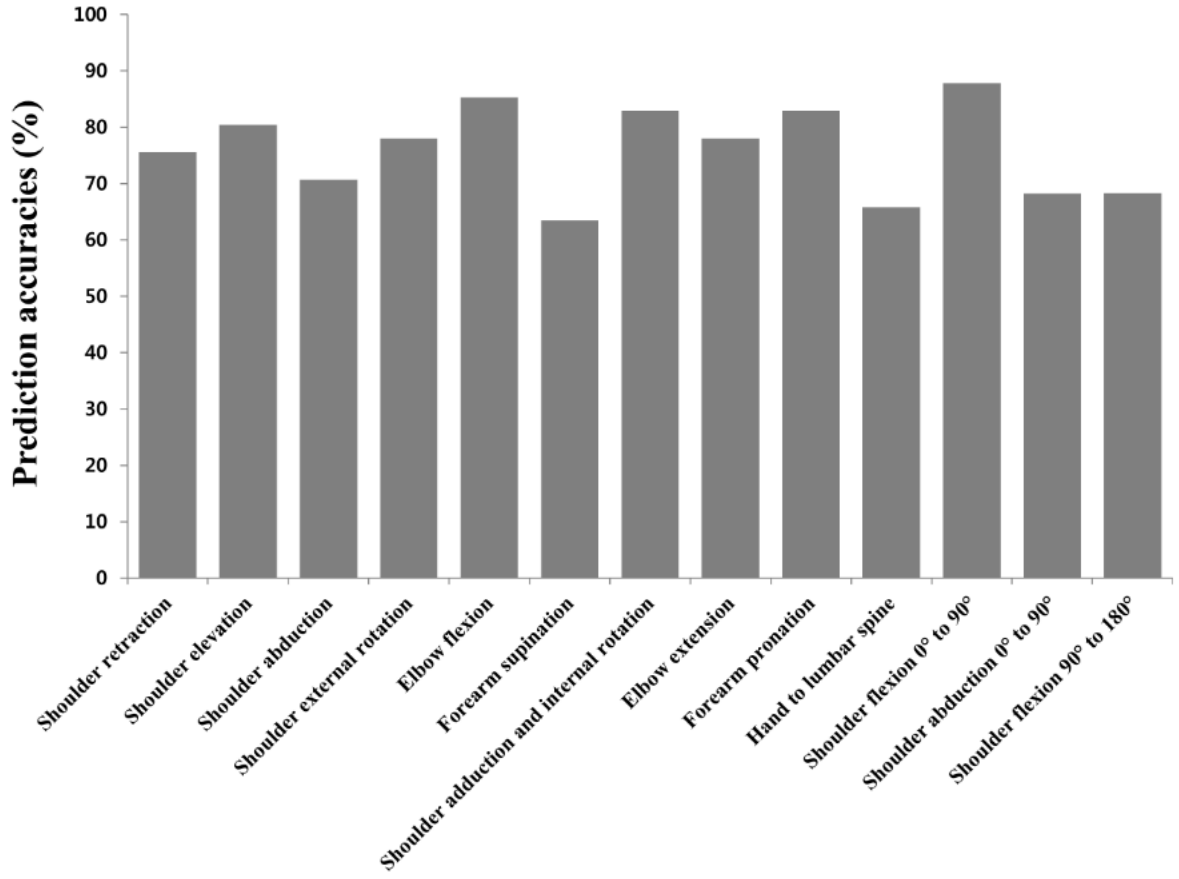


Figure 3.42: Classification accuracies achieved per task of the FMA-UE using a Kinect-based system. Figure taken from the study by Kim et al [19]

conducted by Lee et al was that it required subjects to be moved to and seated in an instrumented room complete with the Kinect sensor and instrumented tools. This clearly required subjects to be mobile or at least to be able to maintain a seated position which is not always possible. Lee et al also claimed that the system reduces the clinician's time requirement to perform the FMA-UE by up to 85% due to system automation of all but 7 of the FMA-UE tasks. There are several flaws in this claim. Firstly, the classifier implemented in the study relied on precise and accurate movements to output a correct score and as such it seems like the presence of a trained clinician would be a requirement during testing. Secondly, the time reduction achieved was based on the 30 minutes required to perform the entire FMA [95] but only the upper extremity section of the FMA is completed in this study. In terms of the classification performance of the system developed by Lee et al, a rule-based classifier was implemented with features extracted based on the guidelines of the FMA-UE. This meant that the study was not limited by the relatively small sample size (9 subjects). Overall, the study showed a very high level of agreement between the clinician and classifier derived scores with only the

tasks of shoulder external rotation (T3) and mass flexion (T18) achieving a percentage agreement of below 70%. The percentage agreement for each task and subtask of the FMA-UE for this study is shown in Figure 3.43. One caveat of the work performed by Lee et al is that testing was performed in a very controlled environment which means that it would be unlikely for these results to be achieved in a normal clinical environment. Another limitation of the study is that the results are only for a very small sample size and therefore may not be representative of tests on a large cohort. The present study was not able to achieve comparably high results but did offer a more robust system in several respects. This includes the use of a much larger cohort of subjects, a wider range of motor deficits due to no requirement for sitting position, and a more robust classification system which can achieve the correct rating even if the task is performed in a slightly incorrect fashion (as is often the case in clinical practice).

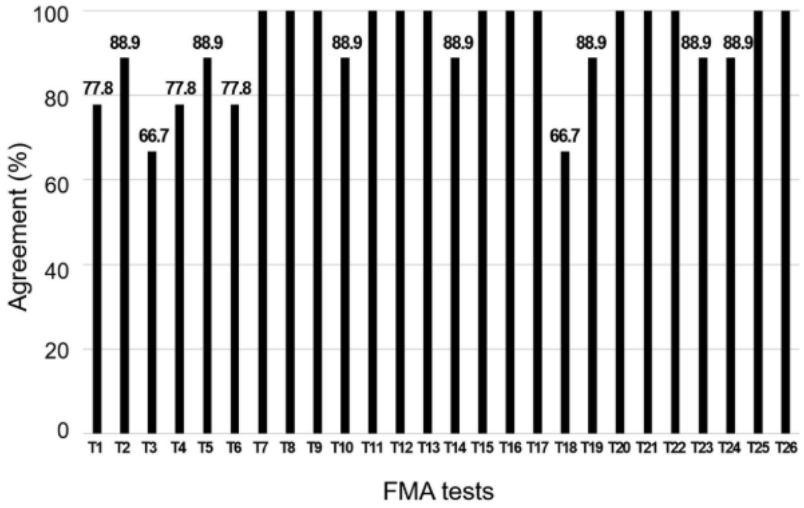


Figure 3.43: Percentage agreement between clinical score and classifier output for 26 task and subtasks from the FMA-UE Figure taken from the study by Lee et al [20]

A final study, which meets the outlined requirements previously stated, has been performed to evaluate an automated system developed Seo et al [67]. Like the prior two studies, the proposed system implemented the Kinect depth camera for motion tracking. Instead of utilising a secondary instrument to measure hand/wrist tasks as in the study by Lee et al, the study instead selected a clinical scale that does not require these movements. The clinical scale chosen was the Mallet classification scale and this is composed of five gross upper arm tasks. Again, similar to the prior study, a custom classification model is developed for predictions to avoid the requirement for significant training data (only

7 subjects tested). The accuracy scores achieved for these tasks ranged from 43% to 100% and averaged 77%. The mean result is comparable to those achieved in the present study (the range is much larger). The results of the work conducted by Seo et al possess limited value due to the small sample size implemented. In addition, the Mallet clinical scale, although useful for classification in terms of only involving gross motor tasks, is not traditionally used in stroke evaluation so has limited clinical relevance.

An overall examination of the literature indicates that the current classification model offers a more robust method of classifying clinical score and with a larger cohort size that most of the prior literature. The classification accuracy results achieved in the present study were comparable with two out of the three studies evaluated. A third study reported higher accuracy results, but these were achieved in very constrained conditions which do not replicate the standard clinical testing environment. Despite the positive results, there are clearly still limitations in the classifier performance which must be examined in more detail.

The classification model proposed in the present study performed less well when classifying a subset of the FMA-UE gross motor tasks. These tasks include those which involved more complex or several stages of movements including the "Coordination-speed" (71%), "Flexor synergy" (75%), and "Hand to lumbar spine" (72%). The high complexity of these tasks means there is a high degree of variety in the way the subjects chose to perform these tasks. The difficulty imposed by these tasks also means that subjects may adopt compensatory movements that are not well predicted by the classifier. The best way to improve classification of these tasks would be to increase the training set, by recruiting more subjects for the study, since this would make the classifier more robust to different movement strategies. It is also possible that there could be better suited kinematic or orientation features to capture these movements which were not implemented in the present study. Another task that was poorly classified (71%), despite being well captured by orientation features, is the "Shoulder flexion (90-180)" task. Presumably in this case there was a degree of interpretation in the scoring which was not well captured by the classifier. For instance, full shoulder flexion may have been achieved by the subject but there may have been the presence of instability or tremor which led the clinician to assign this task a lower score. This would decrease the usefulness of the orientation features in classifying this task since these do not capture these components.

The classification model for predicting hand/wrist tasks performed lower overall (66%) than the same model for predicting gross motor tasks (75%). One reason for this may be

because these tasks involve smaller muscle groups and more subtle movements than gross motor tasks, and as such may be more difficult to classify. Another reason for this finding may be that the MMG derived features used to classify the hand/wrists tasks provide less useful information than the kinematic/orientation derived features used to classify gross motor tasks. Despite this limitation, these tasks are notoriously difficult to classify with inertial sensing so the novel application of MMG to provide a reasonable classification accuracy is still valuable. Prior studies have attempted to classify these tasks with the use of instrumented gloves, but this method has serious sizing and hygiene limitations. Other studies have attempted to use the Kinect sensor to identify these tasks, but this sensing modality does not perform well at detecting these fine motor movements. Overall, the robustness and ease of application suggest that the MMG would be well suited for clinical applications, but more research should be performed to try to improve the accuracy of classification, perhaps through the implementation of new features or fusion of different muscle groups.

Examining the F1-scores across all tasks of the FMA-UE indicates a recurring difficulty classifying subjects who exhibit motor function somewhere between normal and fully impaired (score of 1). One reason for this may be because subjects at this motor function level perform tasks with a wide range of compensatory movements depending on the region of impairment. This would result in tasks being performed with high inter-subject variability which would in turn pose a difficulty in any classification model generalising to new subject data. In contrast, subjects with full impairment (score of 0) or unimpaired (score of 2) would be expected to have lower inter-subject variability due to the absence of and normal movement of the task respectively. Another reason the mid-level motor function rating may have been poorly classified is that for many of the tasks of the FMA this rating may be assigned for a wide range of motor deficits. For instance, many of the tasks of the FMA define a mid-rating as limited movement/ range of motion as compared to not being able to maintain the start position at all (score of 0) or being able to complete the movement fully (score of 2). Limited movement/ range of motion could correspond to a subject having very limited motor control or similarly having just short of full motor control. This clearly introduces a high level of variability in the functional ability of subjects that score 1 as opposed to a score of 0 or 2.

Two-Class Classification

A two class-classification model was proposed as a method of improving upon the aforementioned limitations when classifying the three-classes of the FMA-UE tasks, particularly the low F1-scores attained for mid-level motor function. This was achieved by merging the two classes corresponding to no completion (0) or partial completion (1) of the motor task. This results in two remaining classes that correspond to impaired and healthy motor function respectively. A summary of results for the gross and hand/wrist motor tasks are shown in Table 3.3 and Table 3.4 respectively.

The two-class classification model outperformed the three-class model for both gross motor tasks (85% compared to 75%) and the hand/wrist tasks (74% compared to 66%). This is a moderate improvement but not as much as may be expected given the classification of healthy versus impaired function would be expected to be a much simpler classification problem. The merging of the two classes (0 and 1) has removed the risk of misclassifying between total and partial motor dysfunction but it is apparent there is still substantial misclassification between partial (1) and no motor dysfunction (2) classes. This error is likely due to a combination of the noise in the labels (due to loosely defined classes) and limitations of the classification model (which would be expected to improve with more subjects tested). Despite these drawbacks, the two-class classification model still had a good performance when classifying tasks of the FMA-UE. The resolution of this model is clearly lower than for the three-class problem but still may have applications out with the clinical environment, such as a home-based rehabilitative aid.

3.15.2 Clinical Features

A summary of the correlation results of the clinical features extracted from the data with the local (task) and global (overall) FMA-UE are shown in Table 3.5. Scatter plots showing values of clinical features for each subject versus global FMA-UE scores are shown in Figures 3.35 to 3.41.

All the features extracted showed a positive correlation with the local and global scores. Mean correlation with local rating score was 0.68 (range 0.47 - 0.85) and with global rating score was 0.65 (range 0.56 - 0.78). As a means to contextualise these findings, they can be compared to the results of the study by Fu et al [96], which investigated the concurrent validity of a shortened FMA with other clinical rating scales, also using the Pearson correlation coefficient. This study considered a correlation coefficient of >0.75 to

be excellent, 0.5-0.75 to be good, and 0.25-0.5 to be fair. The shortened FMA was found to have a correlation of 0.57 (good) with the Stroke Impact Scale hand function subscale. On this basis, the correlation coefficients measured between the clinical features and the overall clinical score in the present study all range in the good to excellent range (0.56 - 0.78). This suggests high concurrent validity of the extracted features with the FMA-UE and supports their usefulness as clinical metrics of motor function post-stroke.

The correlation results of this study also compare favourably to the wider literature on automated systems of motor function post-stroke. The study by Julianjatsono et al [69] developed a sensor system consisting of a Kinect sensor and wearable glove. This system was utilised in the development of a regression model of 6 tasks of the FMA-UE. Six features were extracted (one to define each task) and a linear Pearson's Correlation was calculated of the feature compared to the clinical task label for each subject tested. Correlation coefficients were achieved which ranged from 0.17 to 0.475. This is compared to the present study which found correlations of 0.47 to 0.85 when correlating to the same clinical label the feature was extracted from. The study by Julianjatsono et al used a task set to extract features which differs slightly from the present study (one of the tasks used was grasp but otherwise the same tasks were assessed). Despite this, these results give a promising indication of the usefulness of the features extracted in the present study as measures of motor function as compared to the literature.

The clinical features calculated for the present study show great promise both in terms of the correlation results achieved in this study and a comparison with the results of the aforementioned study by Julianjatsono et al. These findings validate the accuracy of the sensor system, the computation of features, and the potential of the clinical feature set calculated. The major caveat to the findings is that features were correlated to a clinical rating scale. The FMA-UE was selected since it forms one of the gold-standard methods of assessing motor function. However, like all clinical rating scales, the output provided is not a measure of "pure" motor function and suffers errors including low resolution and subjectivity. Despite this limitation, the high correlation scores achieved in the present study still show the potential of the clinical feature set.

An examination of the correlation scatter plots as shown in Figures 3.35 to 3.41 enables correlations to be examined in greater depth. Most of the clinical features extracted showed a broadly linear distribution with the global rating score evidencing the potential of these features to be utilised to identify all motor function levels. One exception is the wrist dorsi-volar flexion feature which showed a broadly inverse exponential distribution.

For the wrist dorsi-volar flexion feature it is apparent that at the highest motor function levels, movements are performed with a greater variance of muscle activity. This is unlikely to pose a large problem when classifying since the low- and mid-level motor function levels still appear to follow a broadly linear distribution. In the case of quantifying wrist circumduction using the MMG sensors (Figure 3.41), it is apparent that the variability in response increases proportionally to the FMA-UE total score. This finding does not match the expectation that subjects with a higher motor score would perform the movements in a more predictable way (and as such achieve a more consistent sensor response). A possible reason for this finding is that the subject was not given strict instructions on how quickly to perform the circumduction task (as this instruction is not included in the standard FMA-UE). This may have led to some subjects performing the task with greater speed (and larger muscle activity) which caused a greater variability in response. This should be avoided in any future trials by given more detailed instructions to the subject (such as to perform the task with as much speed as possible).

Inspection of the list of correlation coefficients shown in Table 3.5 may be used to assess some of the limitations of the extracted features. The two features which correlated least well for the FMA-UE global score were the shoulder external-internal rotation (R=0.56) and the wrist flexion-extension muscle activity (R=0.56). The most likely reason for the poor correlation of the shoulder rotation feature is that this movement was captured during the "forearm pronation-supination" task which primarily involves rotation of the elbow rather than the shoulder. Rotation of the shoulder was captured regardless to detect any compensatory movements that the subject may perform, but since these movements are not easily predicted it is reasonable that this feature would not correlate strongly with the FMA-UE global score. The second feature, wrist flexion-extension muscle activity, was captured during the "wrist dorsiflexion/ volar flexion" task and was expected to have been well captured by the MMG sensors. Possible reasons a higher correlation was not achieved include that the feature extracted may not be optimal for MMG sensing or that sensors may have shifted slightly during clinical testing. One feature which performed relatively poorly despite being well captured by the sensor system was the measure of maximum full shoulder abduction (local correlation R=0.66). One reason for this may be because the feature only determines the maximum shoulder abduction present. This means that a high value may be achieved regardless of whether the subject cannot hold the position for any extended time. In addition, the clinician may reduce the score assigned for components not detected by the feature, such as tremor

or instability. This result suggests that while shoulder abduction may be useful as a clinical feature, it is not sufficient in isolation for classification of FMA-UE. Two features which achieved better correlation scores were maximum shoulder flexion (R=0.85) and the range of elbow pronation-supination (R=0.77). These two measures are both relatively straightforward to quantify and indicative of clinical motor function. Shoulder flexion is important in the ability to raise the arm to perform tasks such as writing and drinking. Similarly, elbow pronation-supination is required for grasp and manipulation tasks.

Overall, all the clinical features met the prior specifications (set out in Methods section) of transparency, usefulness, and generability. The transparency and generability have already been discussed and the good level of positive correlation achieved by all the features as compared to the literature demonstrates their usefulness.

CHAPTER

FOUR

Lesion-Symptom Mapping using the FMA and Novel Sensor-based Motor Features

4.1 Chapter Introduction

Numerous LSM studies have been performed to better the scientific understanding of the relationship between motor function and brain anatomy. The state-of-the-art motor outcome measures used for these studies are derived from clinical rating scales, but these suffer limitations including low resolution and subjectivity. The work conducted in this chapter will seek to solve these limitations by the application of a novel set of sensor-derived fine-grained motor outcome measures. A proof-of-concept study will then be performed in the form of a large-scale LSM study.

The outcome measures used for this study are derived from the metrics calculated in the previous chapter. This includes the standard clinical rating scale (to be used as a base measure) as well as the clinical features computed from the automated system. These features encompass angles for each of the joints quantified (shoulder, elbow) and myographic measures of wrist activity. Lesion profiles were manually demarked from a combination of MRI and CT scans collected from the same subjects as recruited in the previous chapter. Lesions maps were extracted from brain lesions which were acute or subacute at the time of motor function evaluation.

The study conducted in the previous chapter illustrated the correlation of the derived clinical features with the FMA clinical rating score and their inherent usefulness as features within a predictive classification model of motor function. Secondly, because these clinical features have a resolution which is only limited by the sensors themselves,

and are independent of the clinical rating scales, it has been theorised that they may offer a more accurate and targetted assessment of "pure" motor function as compared to clinical rating scales. The present study aims to investigate these clinical features further by investigating their usefulness in a large LSM study. More specifically, usefulness of the clinical features beyond that of the FMA alone would be shown in the present study in the form of statistically significant regions detected for subregions of motor function captured by clinical features which are not identified by the respective motor subscales of the FMA.

4.2 Chapter Outcomes

A large scale LSM study is presented in this chapter. The decision was made to use a region of interest type statical analysis to find significant regions. A voxel-based mass univariate approach was considered but then dismissed due to a limited number of suitable subjects and the limitations of this method as set out in the Background chapter. There were two main motor outcome measures used to test statistical significance.

The first motor outcome measure employed in the study is the overall FMA-UE score. This measure was collected from the subjects as part of the study presented in the previous chapter. The FMA-UE was chosen since it forms the main outcome measure used in many prior LSM studies [97, 98, 99] and as such serves as a useful baseline measure to compare to the results found using alternative motor metrics.

The remaining motor outcome measures used in this study are composed of the secondary features calculated in study presented in the previous chapter. These features are calculated directly from the sensor data and as such provide a higher resolution score than possible via the FMA-UE rating scale. Secondly, these features have already been shown to be useful measures of motor function as proven in the previous chapter. The primary goal of this study is to determine whether these features provide greater insights into the present lesion symptom mapping study than the FMA-UE alone.

4.3 Chapter Structure

This chapter structure is as follows:

- **Literature Review**- Review of the prior LSM studies in terms of the primary

methods/ software utilised. The findings of several prior motor-based LSM studies will also be discussed

- **Software**- A description of the software packages and libraries utilised in the present LSM study
- The fundamental stages of the project in terms of:
  - **Subject Recruitment and Imaging**
  - **Lesion Identification**
  - **Lesion Demarcation**
  - **Lesion Map Normalisation**
  - **Lesion Map Transformation**
  - **Analysis**
- **Results**- Statistical results achieved using the prepared lesion maps and motor outcome measures
- **Discussion**- Discussion of the findings in terms of how they relate to previous motor-based lesion-symptom mapping studies, any novel findings, and whether the sensor system provides any added value to LSM studies

4.4 Literature Review

This literature review covers the methods implemented over the fundamental stages of prior LSM studies. This includes the state-of-the-art methodology in addition to any novel processes which have been tested. The main stages of LSM studies through which this review is structured are the detection and demarcation of the lesions, normalisation of the lesion maps, and statistical evaluation of the lesions and outcome measures. Finally, a number of prior motor-based LSM studies are discussed to provide context to any findings achieved in the present study.

4.4.1 Lesion Detection and Demarcation

The current state-of-the-art method for lesion demarcation is by manual tracing. This involves manually demarking the lesion border and is highly time-consuming. In recent years techniques have been developed to reduce this time burden by automated and semi-automated methods.

Manual Methods

Manual methods of lesion demarcation, although time-consuming [100], are still considered the gold standard in terms of accuracy and therefore still implemented in the majority of recent LSM studies. The software chosen for manual lesion tracing is often not stated in the literature. When disclosed, the most frequently selected software cited was MRIcron [97, 101, 102, 103, 104, 5, 105, 106, 107]. Stated reasons for these studies adopting manual tracing methods include that automated methods fail to determine the full extent of the lesion with low image intensity or when there is low contrast between brain tissue and the lesion [108]. There seems to be a general consensus in the literature that although automated lesion demarcation methods have great potential in the future, they are currently not up to the required standard [109]. Once an automated method does become available which offers the same or greater accuracy than manual methods and does not suffer from imaging artefacts, it will be signal a big advance in mitigating the high time costs which currently plagues manual methods. In fact, this time cost is considered to one of the major limiting factors in current large-scale stroke neuroimaging analyses [110].

Automated Methods

The time-intensive nature and variable inter-rater reliability of manual demarcation have led to the development of fully automated methods of lesion demarcation. The stated advantage of these methods is that they require less user interaction, which should reduce demarcation time significantly, and are less observer-dependent than manual methods [21]. A significant disadvantage of many of these methods is that they are less precise than manual demarcation [111], and are inferior at identifying imaging artefacts [21]. Another disadvantage is that automated methods require a very large training dataset [21]. A review by Ito et al [109] found three currently available fully automated lesion demarcation methods. The first of these is the Automated Lesion Identification (ALI) toolbox which uses an unsupervised method that performs outlier detection to segment lesions using a fuzzy c-means algorithm. Secondly, there is the Gaussian naïve Bayes lesion detection method (lesionGnb). This method is a supervised method which uses Gaussian naive Bayes classification for demarcation of stroke lesions. Finally, there is the Lesion Identification with Neighbourhood Data Analysis (LINDA) method. This is a supervised method that uses feature detection and a random forest algorithm to train and classify lesioned voxels.

Semi-Automated Methods

Semi-automated methods have been proposed as a solution to the drawbacks associated with manual and automated methods. This method offers the speed of automated methods with the accuracy and robustness to imaging artefacts of manual tracing. This is achieved by the combination of fully automated lesion detection, followed by manual editing step by the user. The editing steps allows the user to have the final decision on the location and size of the lesion [112]. Haan et al [21] found that the precision of this semi-automated method is comparable to that of manual methods, and can be performed significantly faster than the manual method. An example of a semi-automated method that has been implemented is the Clusterize approach. This method has an automated preprocessing step which involves the identification of the local intensity maxima on each image slice. Clusters of voxels are then assigned based on their relative intensity. A manual cluster selection step is then performed to enable freehand corrections and to optimise the accuracy. An image of the Clusterize approach [21] compared to manual demarcation is shown in Figure 4.1.

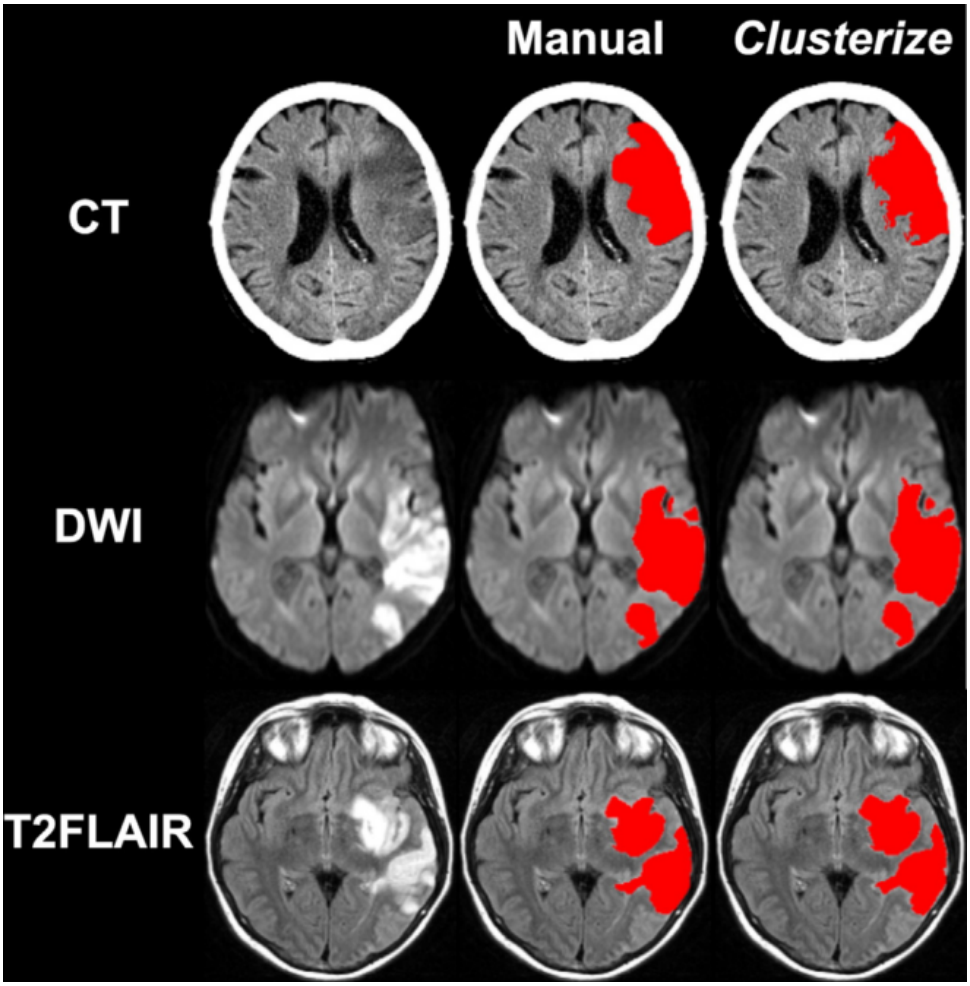


Figure 4.1: Example of the results achieved for using both manual and semi-automated (Clusterize) methods of lesion segmentation. Taken from [21]

4.4.2 Brain Templates and Atlases

A human brain template defines a standard anatomical scan which has been determined from the average of multiple overlaid scans. These have application in analyses which require co-registration or normalisation to a standard template. Brain atlases are determined by similar methods to templates but have been subsequently segmented to provide information of some parameter of interest in the brain. This may correspond to anatomical regions of interest, white matter in the brain, or areas of connectivity. These are of particular use in anatomical exploration and drawing significance from statistical findings.

Both templates and atlases should be defined to the same brain region to ensure consistency of findings between studies. This is achieved by using standardised spaces. Broadly speaking these spaces define how the boundaries of the brain are set based on distance from a given origin. Common spaces which have defined are the Talairach (rarely

used in modern applications) and the Montreal Neurological Institute (MNI) space.

A commonly used template is the International Consortium of Brain Mapping (ICBM) MNI152 NLIN. This defines 152 T1-weighted MRI scans that have been non-linearly registered in the MNI space. This template has found wide application in normalisation functions as discussed later in the section.

Two examples of anatomical region of interest MRI atlases that are currently used are the Atlas of Intrinsic Connectivity of Homotopic Areas (AICHA) [113] and the Automated Anatomical Labelling (AAL) atlas [114]. AICHA is a functional brain region of interest atlas based on the resting-state functional MRI (fMRI) data acquired from 281 subject and normalised to the MNI space. This atlas has been developed for connectivity analyses and provides 192 homotopic region pairs. The AAL brain atlas has a focus on anatomical identification and is normalised in the MNI MRI single-subject brain space. It was originally generated with 45 distinct anatomical volumes of interest per hemisphere although this has now been expanded upon in subsequent iterations. This atlas has been incorporated into popular LSM software (SPM12, NiiStat) for automated lesion mapping analysis. An image of the anatomically segmented AAL3 atlas is shown in Figure 4.2.

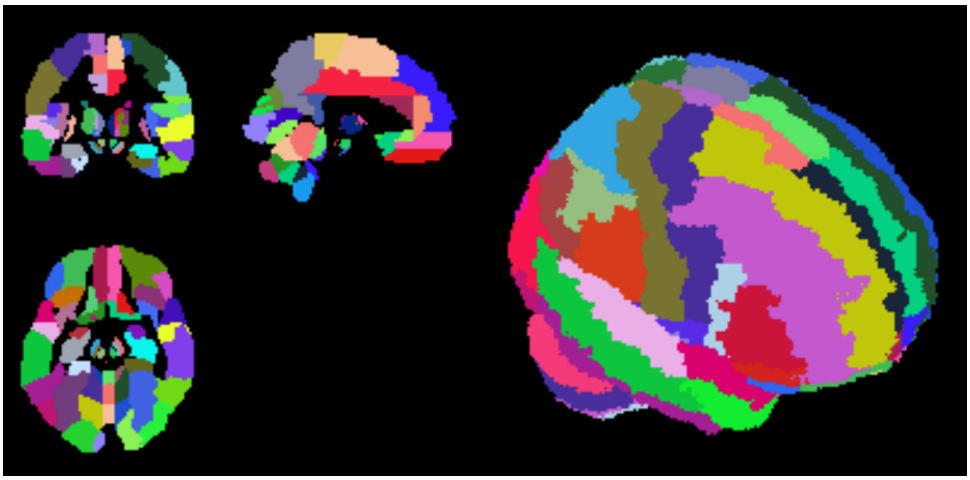


Figure 4.2: Example of the AAL3 atlas in each orientation and as a rendering. Each different colour corresponds to a uniquely identified anatomical region

4.4.3 Normalisation

Normalisation operations are required to translate and transform scans or lesion maps from a local to a normalised template space. This step enables scans to be aligned and

as such is required to make accurate findings that translate beyond single subject trials. One of the most common operations of normalisation prevalent in the literature is the application of functions provided by the SPM12 (Statistical Parametric Mapping 12) toolkit. This toolkit includes a conclusive set of functions for normalisation as well as other aspects of neuroimaging, as discussed in the "Software" section of this report.

The normalise function in the software package SPM12 transforms scans into the standard stereotactic MNI space. The MNI152 NLIN template is implemented as the standard template used to be normalised to during this process. The procedure for the transformation is known as "unified segmentation" for spatial normalisation. This procedure is composed of three steps: segmentation, bias correction, and spatial normalisation. Segmentation refers to the separation of different tissues classes (grey matter, white matter, cerebrospinal fluid). The bias correction is a procedure for removing smoothly varying intensity differences across images. Finally, spatial normalisation is the generation of deformation fields that quantify how the scans should be deformed to transform to the stereotactic MNI space.

4.4.4 Statistical Evaluation

Historical studies of LSM have predominantly been conducted using lesion overlay or subtraction methods. Since the introduction of Region of Interest and later VLSM, these methods have fast become the gold-standard techniques. VLSM has the advantages of higher resolution and no requirement for pre-labelled anatomical regions. Region of Interest methods may be more likely to achieve statistically significant results at lower sample sizes and require less neurological expertise when interpreting results. Common software packages used for statistical analyses in the literature include SPM12 [102], NiiStat [102, 97], the Non-parametric Mapping Software (NPM) [101, 115], and SPSS [116].

4.4.5 Motor-based Lesion Symptom Mapping

The two neural pathways which are understood to be of greatest importance of motor function are the CST and CRST. The origins and pathways of these tracts are relatively well known but their mechanisms are still poorly understood. For this reason, LSM studies have been conducted to try to better understand these mechanisms and the relative importance of different segments of these pathways for motor function and recovery. A

subset of relevant motor based LSM studies will be discussed here.

Early brain mapping studies of damage to motor function sought to examine the relationship with lesion size only (relative or absolute size of brain lesion). The 2000 study by Chen et al [117] found no or only a weak relationship between relative/ absolute lesion size and whole body motor outcome/ recovery. A subsequent 2001 study by Binkofski et al [118] found that the initial lesion size could be used as a predictor of initial hand motor score, but that this did not hold true for long term motor recovery. The overall conclusion of these studies is that lesion size alone is a poor predictor of motor function. This is due to compensatory ability of the brain being much more limited at certain critical locations. Even small lesions to these locations have potentially large detrimental effects on motor function. This has led to the emergence of studies that focus on the lesion profile (a combination of size and location) rather than the size alone.

Subsequent studies have incorporated the location of the lesion in the analysis as performed in RLSM and VLSM studies. An early 2001 study by Shelton et al [119] observed the highest probability of upper extremity motor recovery associated with isolated lesions in the cortex, followed by the sub-cortical regions of the corona radiata, and the internal capsule. This is supported by the 2008 study by Schiemanck et al [120] which found that lesions of cortex or other sub-cortical regions had a significantly higher probability of isolated hand motor function as compared to lesions of the internal capsule. The 2010 study by Lo et al [121] identified the junction of the corona radiate leading into the CST as being the most detrimental region of damage for upper extremity motor function. The study by Chen et al [122] found that damage to the regions of corona radiata, internal capsule, and insula all had a statistically significant effect on whole body motor function as assessed by the modified Rankin Scale. A similar study by Wu et al [123] found statistical significance of worse motor outcomes in the same regions in addition to the external capsule, superior longitudinal fasciculus, uncinate fasciculus, postcentral gyrus, putamen, and operculum. The internal capsule has also been indicated to have a significant role in poor long term upper extremity motor function as shown in a study by Lee et al [99]

One conclusion which may be gathered from all these studies is that cortical lesions do not appear to lead to the worst motor outcomes. In fact, there is evidence that these regions may be well compensated for by spared motor cortex areas, such as the pre-motor cortex [124]. Another conclusion that may be drawn from the aforementioned studies is that sub-cortical regions that are present in the CST, such as the internal capsule, corona radiate, midbrain, pons, or medulla, all result in worse motor outcomes. This is likely

due to several factors including their relative importance of these regions in the CST, small size or the limited compensatory ability of surrounding tissues.

Several studies have directly studied the relationship between motor function and damage to the CST and CRST pathways. These studies have centred around reinforcing and improving existing knowledge that the CST has a primary role in distal extremities (including hand function) whereas the CRST has a role in axial muscles (including locomotion). The 2010 study by Zhu et al [41] found that upper extremity motor recovery was inversely proportional to the lesion load of the CST. The influence of CST damage has also been shown to predict the development of hand spasticity in a study by Plantin et al [97]. A 2014 comparison study by Yoo et al [125] found that subjects who exhibited damage to the CST did indeed show worse hand function than those with damage to the CRST. The inverse was also found to be true in terms of gait function. This is supported by the study by Jang et al [126] which found that the increased fibre volume of the CRST had a significant effect on chronic stroke subject's ability to walk.

4.5 Software

The software utilised in the present study was chosen from established software packages that have been widely implemented in previous LSM studies. These packages may be broadly categorised by the stages of imaging and lesion mapping (MRIcron), transformation and normalisation (SPM12), and statical analysis (NiiStat).

4.5.1 MRIcron

MRIcron is widely used in the neuroimaging field for the visualisation of scans. It provides functionality for in-depth navigation at the voxel level for each scan plane. Another commonly used application is the ability to define custom overlays on top of the scans. This is frequently used by studies as a means of manually demarking lesions or other features of interest. Additionally, existing overlays or statistical lesion maps may be added to scans as a well of interpreting the significance of results. A final common application of MRIcron is the preparation of scans prior to analysis. This is provided by functionality to convert scans from their native format (.DICOM) to the more usable .NIfTI format. An example of the MRIcron GUI is shown in Figure 4.3.

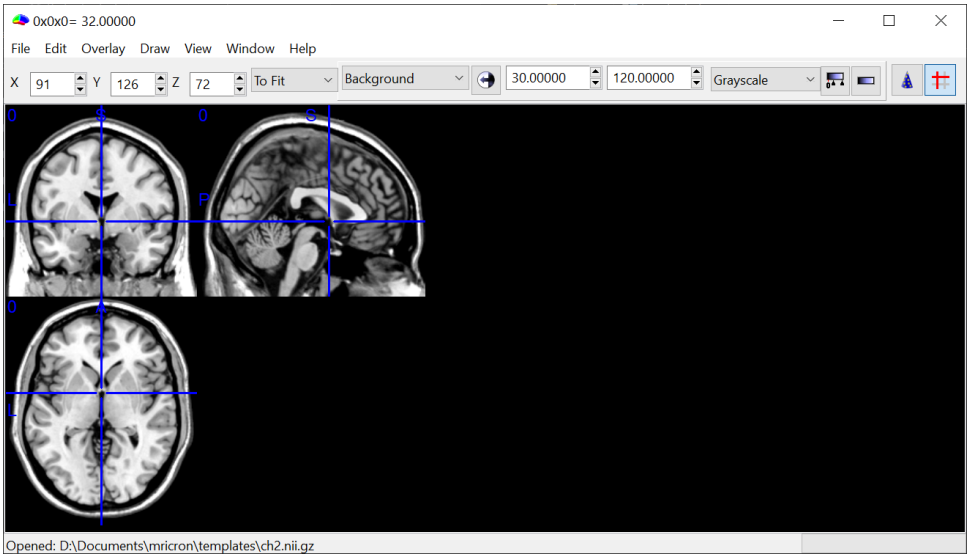


Figure 4.3: GUI for the MRIcron software

4.5.2 SPM12

SPM12 is the latest of a series of SPM software packages developed for the analysis of structural and functional imaging data. It was developed by the Wellcome Department of Imaging Neuroscience (University College London) as a toolkit within Matlab. The GUI provides a wide variety of functionality for the assessment of neuroimaging data. A subset of these functions required for the present study are as follows. The "Display" function provided visualisation and manual adjustments (origin and transformation) to be applied to individual scans. The "Check Reg" command enabled comparisons of multiple scans to check the performance of any transformation operations. Finally, the "Coregister" and "Normalisation" commands were used for these respective operations in the present study. An image of the SPM12 GUI is shown in Figure 4.4.

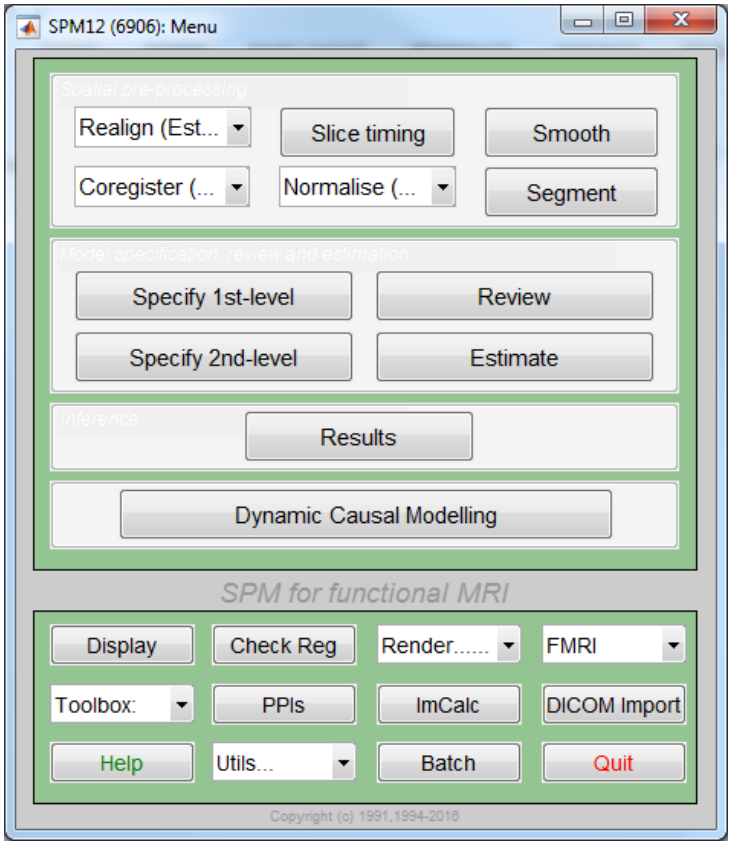


Figure 4.4: GUI for the SPM12 software package

4.5.3 NiiStat

NiiStat is a Matlab toolkit that has been developed for the analysis of neuroimaging data from clinical populations. Statical analyses which may be performed include voxel and region of interest-based statistical testing. The toolkit requires lesion maps to be drawn in the MNI space and converted into ".mat" format files. Design files should be formatted in Excel and these provide information on which lesion maps correspond to which outcome measures. An image of the NiiStat GUI window is shown in Figure 4.5.

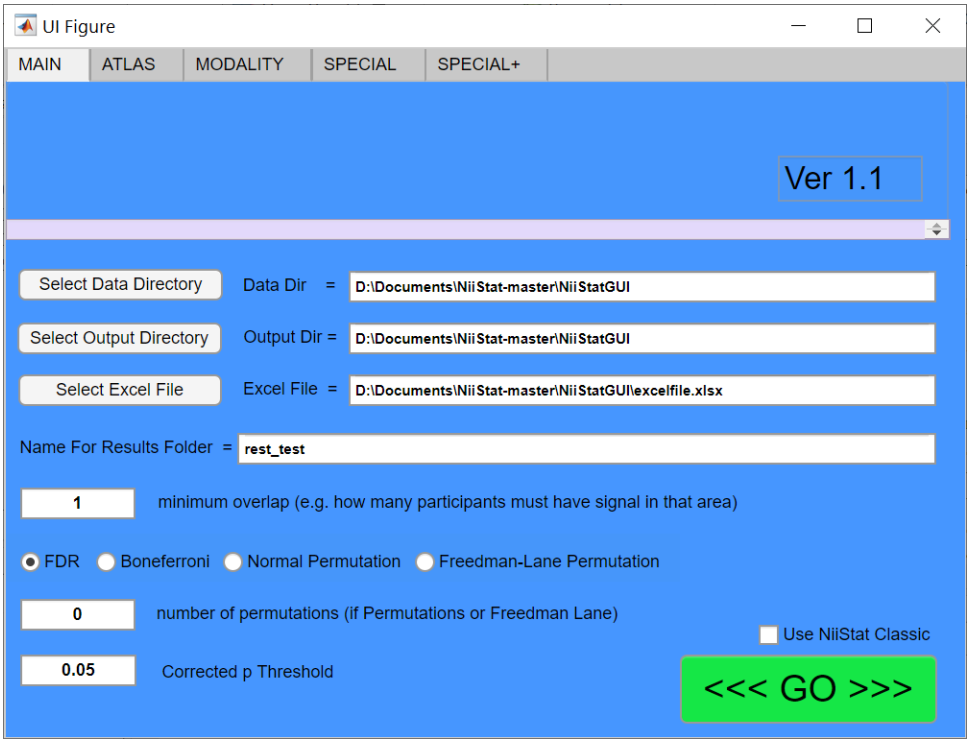


Figure 4.5: GUI for the NiiStat Toolkit

4.6 Subject Recruitment and Imaging

The same cohort of subjects as recruited for the study detailed in the previous chapter was also recruited for the present study. This cohort consists of 64 subjects recruited from the acute and hyper-acute wards of Charing Cross Hospital (London). There were no subject requirements in terms of a minimum level of required motor function. Cognitive ability was assessed before clinical testing by the administration of the GCS. A full score was required prior to testing to ensure the subject was capable of giving informed consent

to participate in the study and was able to follow verbal commands.

CT and MRI scans were already administered to the vast majority of subjects as part of their standard post-stroke treatment. These scans were subsequently gathered from the hospital repository system with the subject's permission. In the case of two subjects, existing scans were not suitable for this study so additional scans were administered with the subject's express consent. The vast majority of scans available for subjects were taken in the acute/ subacute phase, imaging phases which are defined for the present study in Table 2.2 of the Background section. A full breakdown of each potential subject including the neuroimaging modality used is provided in Appendix A Table 8.2.

On clinical review of scans, a total of 13 subjects were found to present with no discernible acute or subacute lesions and were therefore removed from this study. This left a remaining 51 subjects available to participate in the study. A total of 41 of the remaining subjects were imaged with MRI scans while 10 subjects were imaged with CT. The MRI scanner used for scanning was either the Siemens Avanto 1.5T or Siemens Verio 3T. Scans were selected either from the FLAIR or DWI sequences for further processing. All CT scanning was performed using the Siemens Definition 128 slice scanner.

As discussed in the Chapter Outcomes section, the motor features utilised in this study were taken directly from the study discussed in the previous chapter. These subjects were all assessed on the acute and hyper-acute wards of Charing Cross Hospital. Subjects were admitted to these wards immediately after the onset of stroke and only present on these wards for a limited period of time. For this reason, all subjects assessed for this study were considered to be within the acute and sub-acute stroke phases, as defined in Table 2.2. The only exception to this were subjects who were admitted with stroke-like symptoms but could not be considered to be in the acute/ subacute phase since they presented with no lesion or only chronic lesions on any scans. These subjects were already excluded from the study on the basis of not presenting with an acute or subacute lesion when imaged.

Scans were originally recorded in the native ".DICOM" format (Digital Imaging and Communications in Medicine). All files were later converted into the ".NIfTI" format, using the MRIcron software, for ease of subsequent operations.

4.7 Lesion Identification

Most subjects had several scans available in addition to the one taken immediately post-stroke. These scans had been taken either historically or as part of their follow-up care. A

single, primary scan was selected for each subject which would later be used for drawing the lesion maps. This primary scan was selected as the one taken in the acute or subacute stroke phase. If both MRI and CT scans had both been taken during this period, the MRI was selected over the CT scan due to the improved resolution offered by this modality. All subsequent lesion demarcation was performed on the primary scan although secondary or historic scans were also used as references to identify any previous lesions.

The clinical notes made by neurologists at Charing Cross Hospital were also available for this study and used to assist in the lesion identification and demarcation stages. These notes were particularly useful in distinguishing between acute and chronic lesions (when both were present), and for identifying the appropriate lesion boundaries.

4.8 Lesion Demarcation

The manual method of lesion demarcation was selected for this study. This method was chosen since it remains the gold standard demarcation method. In addition, different modalities and sequences of scans were present in the study which may have led to the inclusion of imaging artefacts if using automated demarcation methods. Secondly, several subject scans also included the presence of chronic lesions due to prior stroke. The use of manual lesion demarcation ensured that the chronic lesions were not included in any subsequent lesion maps.

The software MRIcron was selected for lesion demarcation. This software was selected due to its widespread application in many prior LSM studies, as discussed in the literature review. MRIcron also provides several useful features that makes the demarcation more straight-forward. These include easy navigation of the different scan slices, ease of drawing and filling the boundaries of drawn lesions, and the use of overlays to compare different boundary maps that have been drawn.

Lesion demarcation was performed by the author after being trained on the procedure by a consultant neurologist at Charing Cross Hospital. The lesion boundaries drawn for the subjects were also verified by the consultant neurologist and assistance provided for any cases where the lesion boundary could be considered ambiguous. Lesions were demarked from the primary scan of each subject tested. This involved manually drawing a boundary around each of the acute/ subacute lesions and then selecting the area as a lesion for every slice available. An example of the use of MRIcron in the present study to draw a lesion boundary for a single MRI scan slice is shown in Figure 4.6.

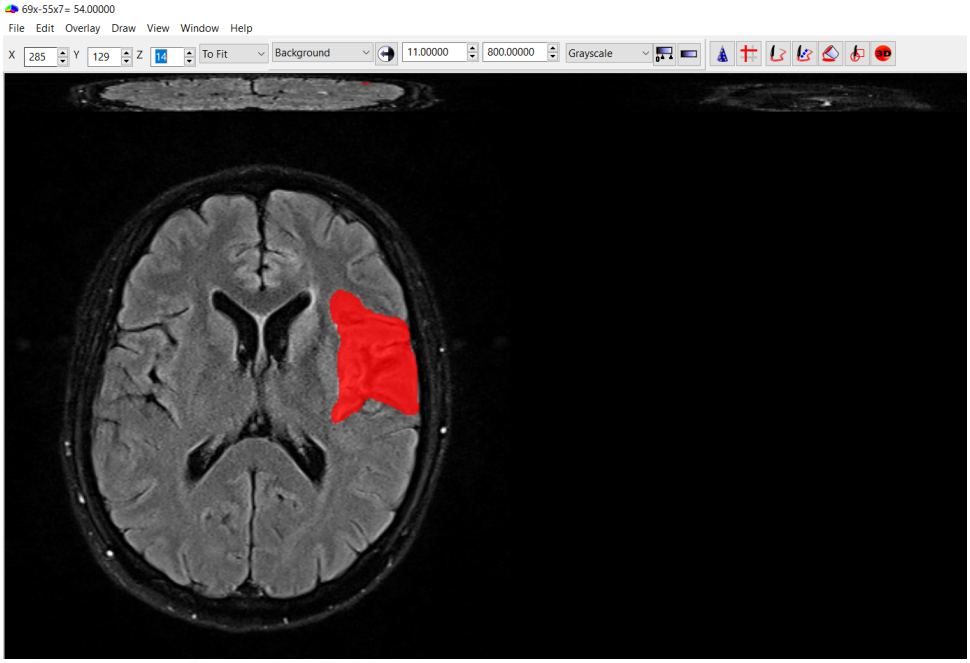


Figure 4.6: Example of lesion boundary drawn in a single slice using the MRIcron software

4.9 Lesion Map Normalisation

The software package SPM12 was selected for all normalisation operations in the present study due to its simple interface and the powerful functions provided. The major normalisation steps included in this study were manual transformation, normalisation, and verification.

Manual transformation methods are recommended to help ensure that subsequent normalisation operations achieve the optimal transformation, as opposed to transforming to a local minimum. For the present study, this involved the reorientation and modification of the origin of the primary scan. Operations were performed using the "Display" function on SPM12. This function opens a new window which provides buttons for simple transformation operations, magnification, and setting of the range of intensity values (see Figure 4.8). The origin was to be set at the approximate location of anterior commissure (AC) to best match the template scan. This was achieved with the aid of images of the AC taken from standardised scans with different magnification levels (see Figure 4.7). Once the appropriate transformation values were found, these were applied to the primary scan and the previously extracted lesion map.

Normalisation was performed in SPM12 using the latest normalisation function. This operation involved deriving the transform function necessary to transform the primary scan of each subject to MNI152 NLIN template. Once this function had been derived it

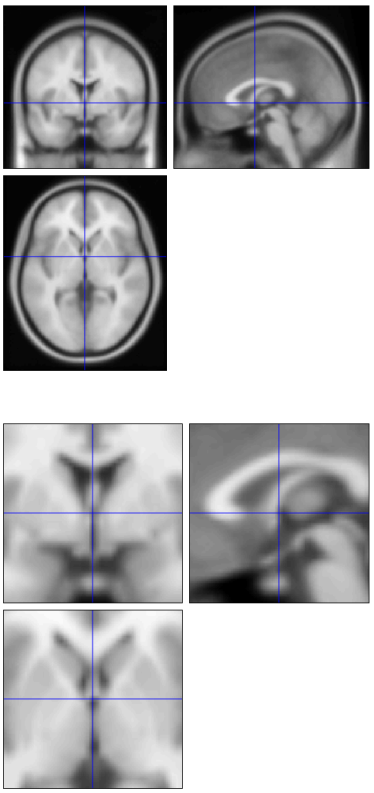


Figure 4.7: Template reference images taken from the 15 subject T1 scans in the MNI reference frame (SPM12) and set at different magnification levels

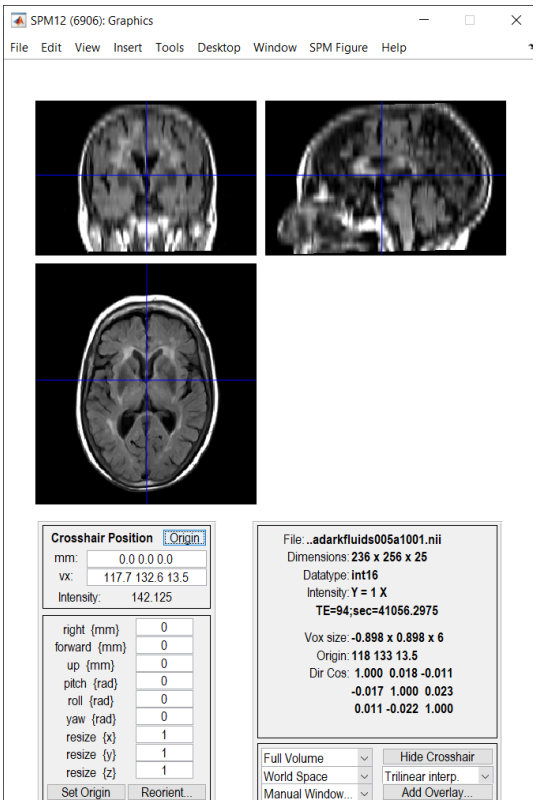


Figure 4.8: SPM12- Display function for manually reorienting and setting the origin of scans

could be applied to primary scan and lesion map. More in-depth details of the stages of normalisation in SPM12 are given in the Literature Review.

The default SPM12 parameters were selected for this operation (see Figure 4.9) since these were found to perform well for all the scans of the study. Two parameters worth mentioning are the voxel size of the resulting scans and the interpolation method. A voxel size of 2*2*2 mm was selected for normalisation since this was considered to be a good compromise between the accuracy of lesion maps (limitations of human error considered to be higher than this resolution) and size of the resulting dataset. The interpolation method 4th degree B-spline interpolation was selected. This method is slower than trilinear interpolation, but the computation time of this stage was not considered a major limitation for the present study.

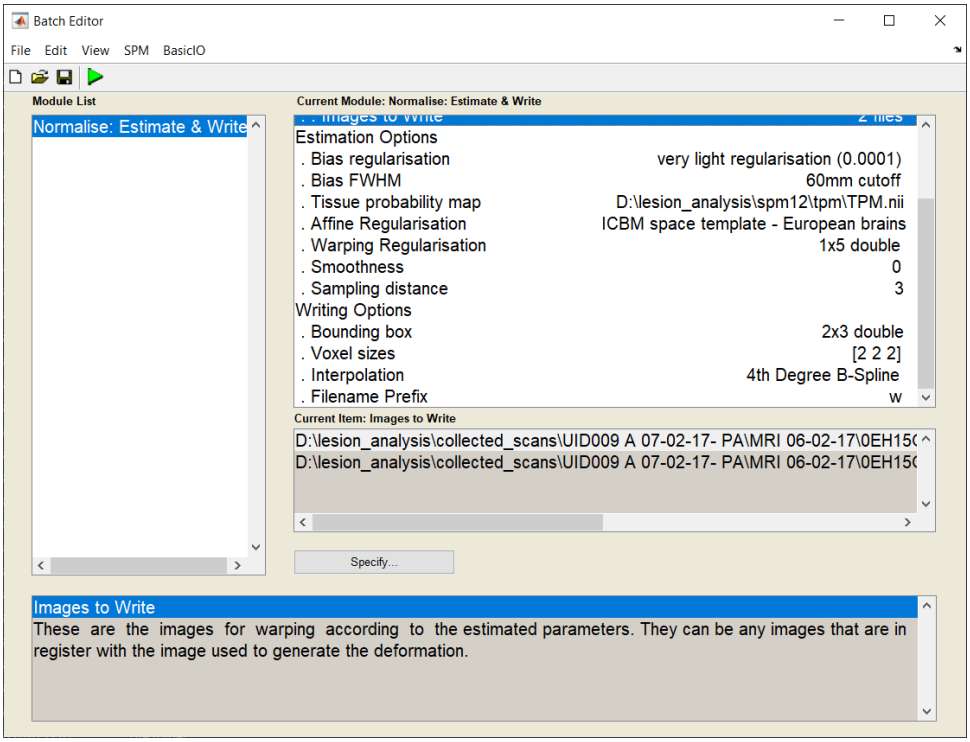


Figure 4.9: SPM12- New Normalisation Function (default parameters)

Finally, a validation step was performed to ensure that normalisation operations had been completed successfully for each scan. The "Display" function was once again used for visualising the orientation and origin of each scan. This was performed for each scan individually to ensure there were no results that deviated from the norm. Next, MRIcron was used to overlay all the normalised scans on top of a 15 subject T1 scan MNI space template image. This was used to ensure that all the scans fit the same boundaries and alignment. Since the lesion maps were already co-registered to the primary scans, there was no need to additionally validate the transformation of these maps.

4.10 Lesion Map Transformation

This section refers to transformations that were applied to the lesion maps post-normalisation. For the present study, the decision was made to flip lesion maps so that lesions were predominantly present on the left side only. This is a frequent practice in motor function based LSM studies [127, 115, 97] which takes advantage of the symmetrical nature of motor pathways in the brain. By flipping lesions onto one side, the statistical analysis of subsequent operations may be drastically improved. For the present study, each le-

sion map was inspected individually and flipped, if necessary, so that the majority of the lesions were on the left-hand side. The decision was made to flip onto the left-hand side since this is where most of the subject lesions were already present. A heat-map of the normalised lesion maps projected onto the standard single subject T1 MNI space template is shown in Figure 4.10.

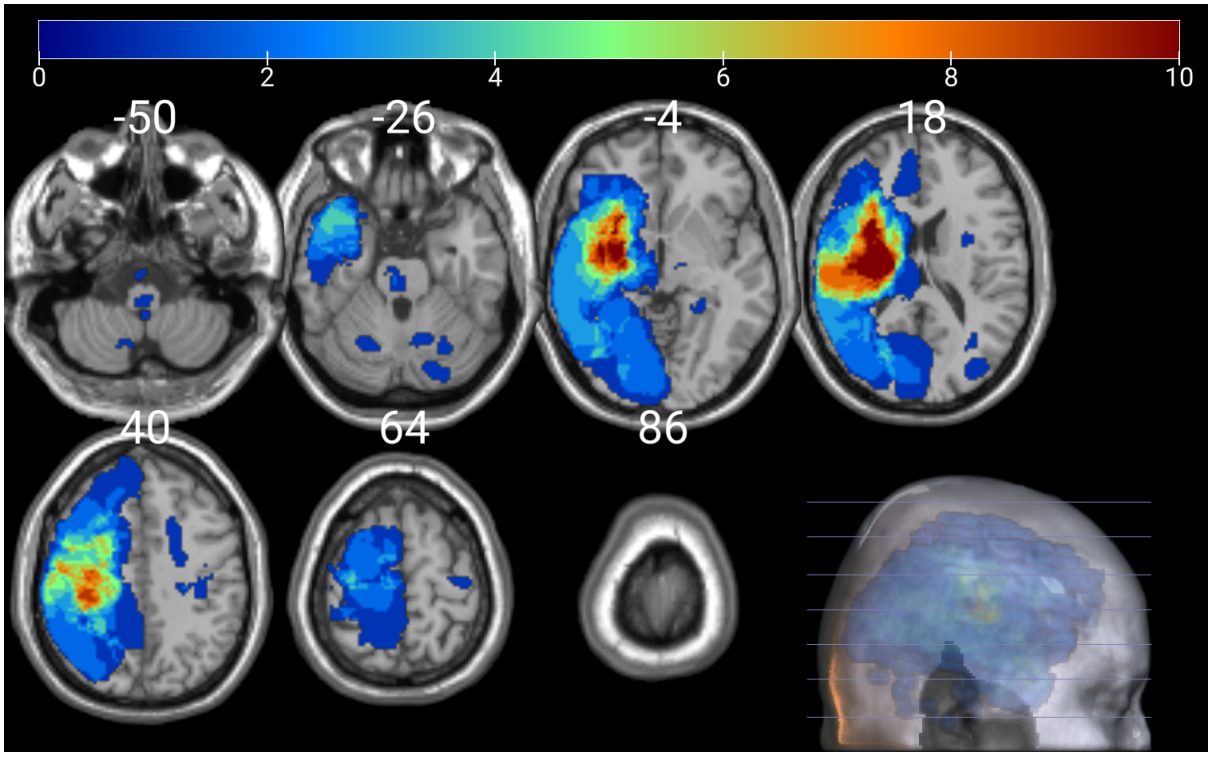


Figure 4.10: Heat-map of the lesions from all subjects imaged who met the study criteria (lesions flipped onto the left-hand side). A scale is present which represents the least (dark blue) and most (dark red) commonly present lesions

4.11 Analysis

The statical portion of this study was performed using the NiiStat statistical software due to its widespread application in prior LSM studies. A total of eight independent analyses were selected to be performed. One of these analyses involved the overall FMA-UE score as assigned by the clinician. The findings of this analysis were to be used as a base measure for all subsequent results. The remaining seven analyses each involve one of the sensor-derived clinical features (see Table 3.5) as the predictor. Since each of the clinical features corresponds to a subcomponent of motor function it was hoped that

these outcome measures would uncover significant findings beyond the baseline measure.

Design files were produced in Excel corresponding to each analysis. These were set up with a single predictor corresponding to the motor outcome, as well as reference to the lesion maps of interest. All the predictor variables were considered to be discrete and the lesion maps binary (since lesions were drawn as present or not). Permutation thresholding was selected as the analysis method since this is the gold standard method in LSM analyses for controlling family-wise error [128]. This method involves running many iterations of random orders of outcome measures to detect whether observed effects are unexpected. For this study permutation testing with 3000 permutations was selected. In addition, a corrected P threshold of 0.05 was used to identify statistical significance. Finally, only damaged voxels which that present in at least two subjects were included to avoid false positives in lesions only present in one subject. The settings chosen for the analysis are shown in NiiStat GUI as shown in Figure 4.11.

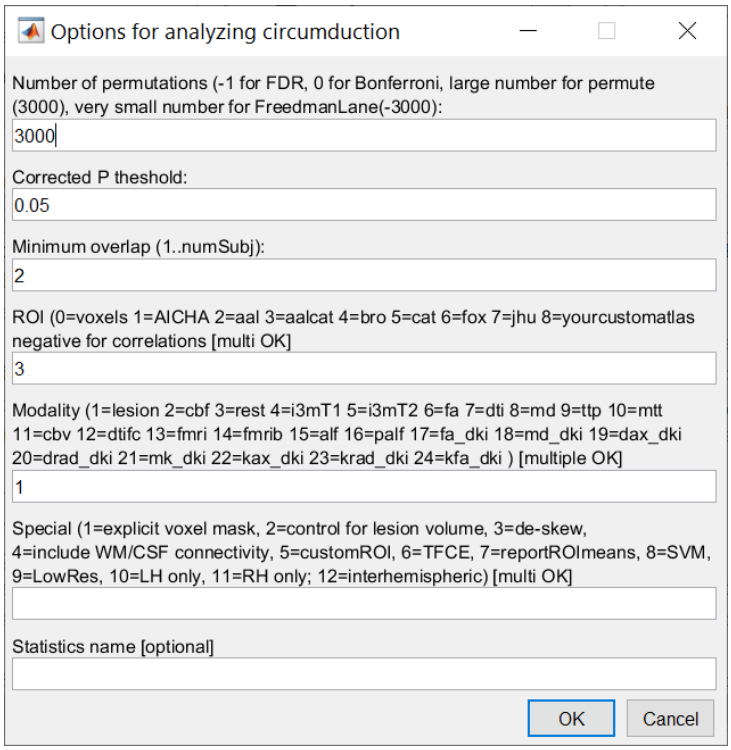


Figure 4.11: GUI for the NiiStat Statistical Analysis

A region-of-interest type of analysis was selected. This was chosen due to the relatively low number of subjects once exclusions had been considered. Region of interest analyses are also widely used in the literature [102, 129], and have useful properties including high statistical power and ease of subsequent analysis. The region of interest analysis was com-

bined with the AAL atlas, to provide all potential regions where statical significance may be achieved. This atlas was chosen due to the detailed level of anatomical segmentation offered [130].

4.12 Results

This subsection will detail the results of the large-scale region of interest analysis performed in the present study. Statistical significance was assessed via permutation testing using two groups of motor outcomes. Firstly, the overall and subsections of the FMA-UE were tested for significance to use as baseline results. The statistically significant regions which were determined for these measures are shown in Table 4.1. Secondly, statistical significance was tested for seven clinical features and these results are presented in Table 4.2. No statistical significance was found for the shoulder twist rotation, shoulder abduction, and elbow magnitude clinical features and as such these features do not have any regions listed within the table.

For the clinical features which achieved statistical significance, the significant regions overlaid onto the ALL atlas are shown in Figures 4.12 to 4.15. The keys on this figures represent the z-value and the minimum z-value required for $p < 0.05$ significance is included within the figure caption. For overlays that contain multiple statistically significant regions, the darkest colour represents the minimum threshold for significance while the brightest colour represents the highest level of significance achieved of all the regions.

Table 4.1: The statistically significant regions on the ALL atlas as detected using permutation testing with the overall and subsets of the FMA-UE rating scores for each subject. Statistical significance determined by permutation testing with a corrected threshold of $p < 0.05$ (one tailed significance test) required for significance

FMA-UE Score	Statistically significant regions
Overall score	Corticospinal tract, internal capsule
Upper extremity subscale	Corticospinal tract, internal capsule
Wrist subscale	Corticospinal tract, internal capsule
Hand subscale	Corticospinal tract

Table 4.2: The statistically significant regions on the ALL atlas as detected using permutation testing with the clinical features derived for each subject. Statistical significance determined by permutation testing with a corrected threshold of $p < 0.05$ (one tailed significance test) required for significance

Clinical Feature	Statistically significant regions
Shoulder int-ext rotation (range)	None
Shoulder abduction angle (max)	None
Shoulder flexion angle (max)	Pallidum
Elbow pronation-sup (range)	Putamen, corticospinal tract, internal capsule
Elbow flexion (range)	None
Wrist flex-ext muscle activity	Cortico-ponto-cerebellar tract, corticospinal tract, internal capsule
Wrist circumduction muscle activity	Corticospinal tract

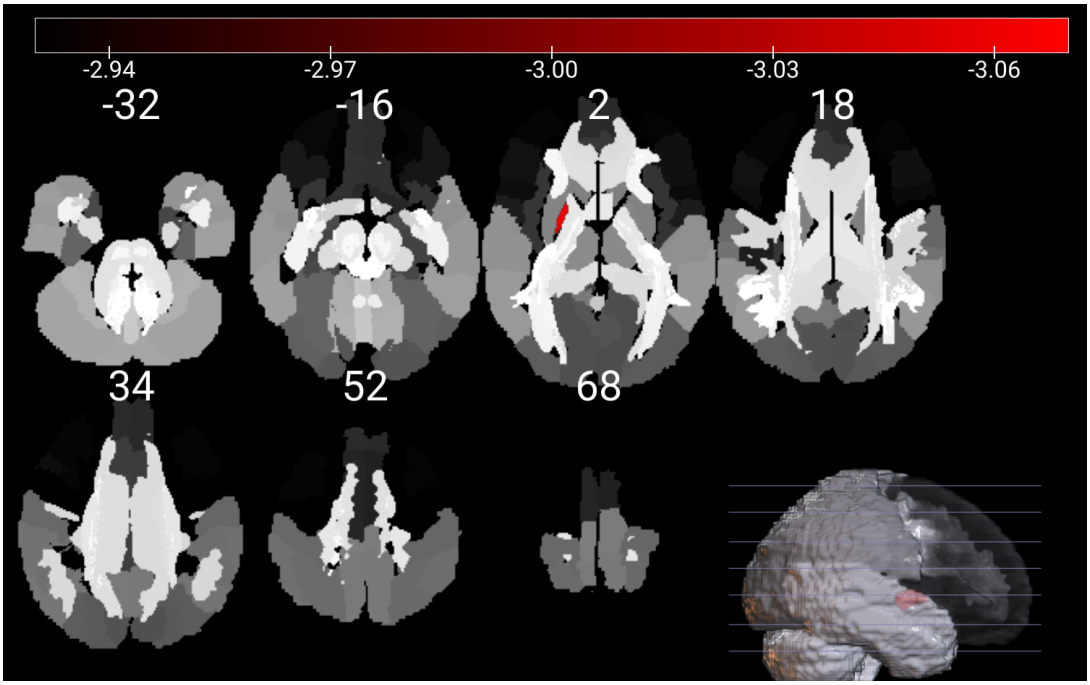


Figure 4.12: Region-of-Interest results for the maximum shoulder flexion angle outcome measure (pallidum found to be statistically significant and is highlighted in the figure). Key at the top of the figure represents the z-score ($z < -2.94$ represents $p < 0.05$ statistical significance)

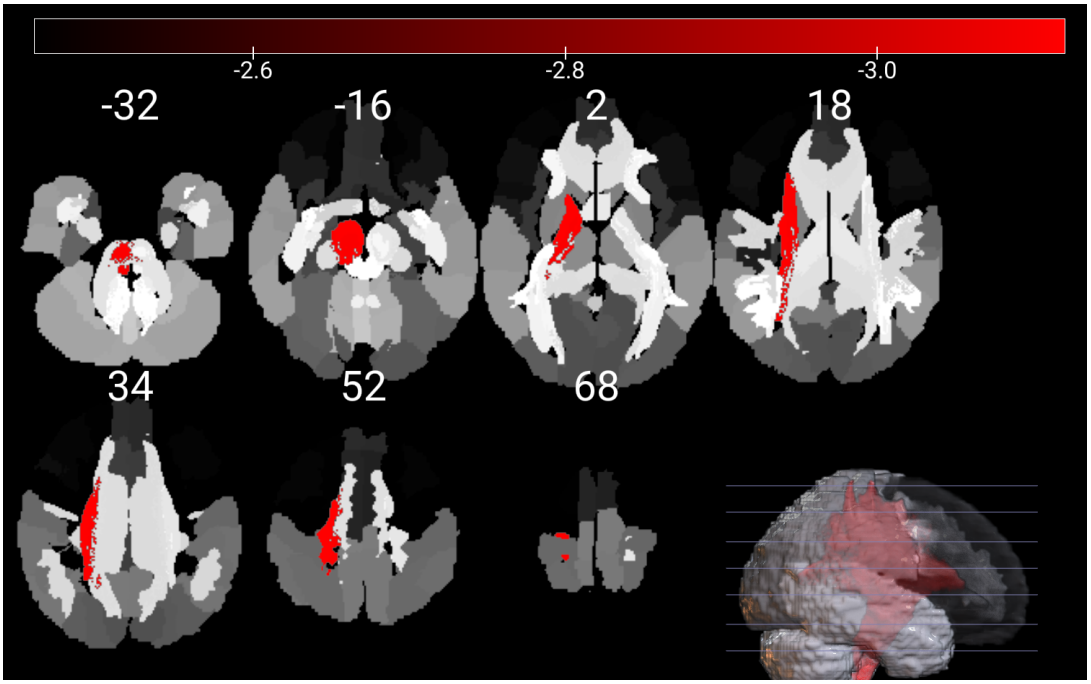


Figure 4.13: Region-of-Interest results for the range of elbow pronation-supination outcome measure (putamen, corticospinal tract, and internal capsule found to be statistically significant and are highlighted in the figure). Key at the top of the figure represents the z-score ($z < -2.46$ represents $p < 0.05$ statistical significance)

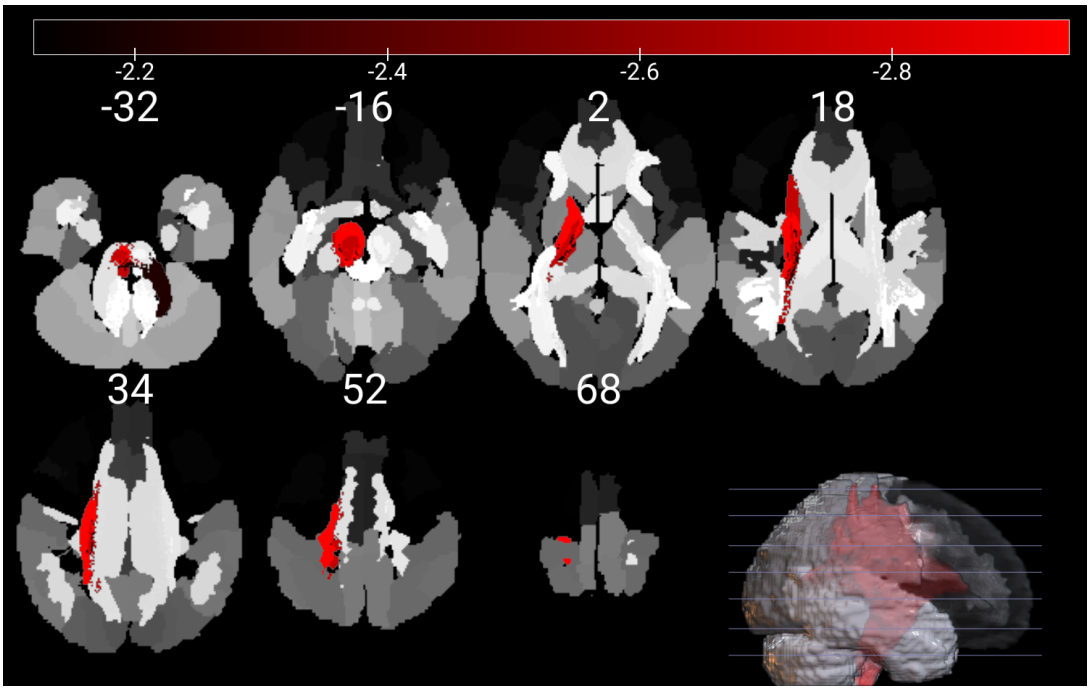


Figure 4.14: Region-of-Interest results for the wrist flexion-extension muscle activity outcome measure (cortico-ponto-cerebellar tract, corticospinal tract, and internal capsule found to be statistically significant and are highlighted in the figure). Key at the top of the figure represents the z-score ($z < -2.12$ represents $p < 0.05$ statistical significance)

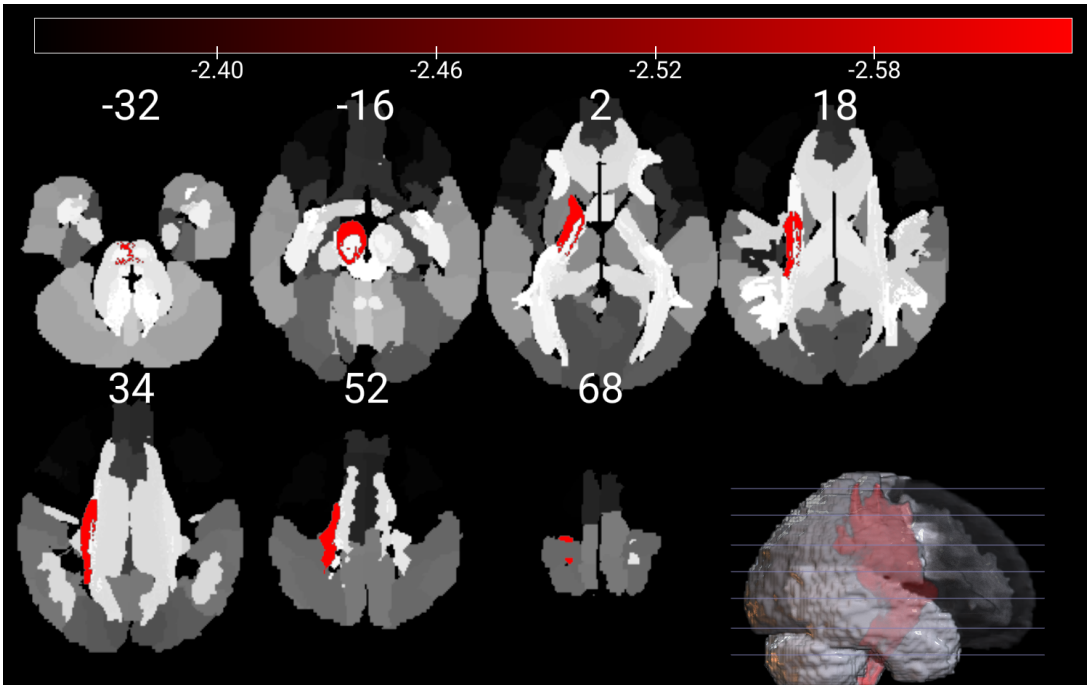


Figure 4.15: Region-of-Interest results for the wrist circumduction muscle activity outcome measure (corticospinal tract found to be statistically significant and is highlighted in the figure). Key at the top of the figure represents the z-score ($z < -2.35$ represents $p < 0.05$ statistical significance)

4.13 Discussion

For this section, a summary of findings of the present LSM study will be discussed with respect to how they relate to current anatomical knowledge of motor regions and pathways of the brain. Next, a comparison will be made with the findings of prior motor-based LSM studies and the significance of any novel findings discussed. Finally, any limitations of the present LSM will be discussed in more depth.

4.13.1 Statistical Results

The clinical features implemented in this study may be broadly considered to represent subcomponents within two different forms of motor function: fine and gross. The myographic clinical features (flexion-extension and circumduction) describe wrist motor function. Since the majority of prior LSM studies present results in terms of fine or gross motor function we will consider these wrist features to be representative of fine motor function to enable direct comparison with the literature. These movements may not perfectly conform to the clinical definition of fine movements, but we make this assumption since they fit this definition much more than they do gross motor function. The remaining clinical features may be considered to measure subcomponents of gross motor function.

The first results to be discussed are the LSM statistical findings when selecting the overall and subscales of the FMA-UE as the outcome measures. The use of a clinical rating score is the prevalent method used in prior LSM so any findings from this analysis are important to discuss as a baseline. Next, the statistical results of the clinical features will be discussed both in terms of the anatomical relevance and in terms of whether these measures provide a greater depth of findings than possible using the rating scales alone.

FMA-UE Rating Scales

Reviewing the results of permutation testing using the FMA-UE overall and subscales shown in Table 4.1 shows that only the regions of corticospinal tract and internal capture showed statistical significance. Both of these regions were also found to be significant for tasks more closely associated with gross motor function (upper extremity subscale) and fine motor function (wrist subscale).

The primary role of the CST is discussed in the Background section and is considered to have a major function within the control of the distal extremities (particularly within

fine motor control). However, the CST and CRST follow a similar pathway and damage to one rarely occurs in isolation. Disruption to the CRST is associated with deficits to gross motor function. For this reason, the finding that damage to the CST (and CRST by proxy) could result in a motor deficit to the upper extremity, wrist, and hand subscales is an expected finding.

The internal capsule was also found to be statistically significant for all subscales except the hand. The internal capsule is a relatively small region and a large part of it is dedicated to the CST, carrying motor information from the primary motor to the lower motor neurons in the spinal cord. This supports the finding that damage to this region may have a large consequence on both gross and fine motor function.

Clinical Features- Components of Gross Motor Function

The clinical features representing subcomponents of gross motor function achieved significance with several regions of the brain and these results are represented in Table 4.2 and Figures 4.12 and 4.13. Broadly speaking the regions which were found to be significant for these clinical features correspond with the current functional understanding of the brain. These results are discussed in more depth below.

The CST region was statistically significant for the elbow pronation-supination secondary feature, a movement which forms a major component of many gross motor actions. The shared pathway of the CRST and CST has already been discussed (and the role of CRST in gross motor actions) and therefore this finding aligns with the existing knowledge of this region.

The pallidum was found to be statistically significant for the only the shoulder flexion clinical feature. The pallidum is composed of two regions: the global pallidus and the ventral pallidum. The global pallidus is understood to have a major role in the motor system, particularly within the regulation of voluntary movement [131]. The pallidum is also thought to have a primarily inhibitory action that acts in tandem with the excitatory action of the cerebellum. Damage to the pallidum has been found to result in movement disorder characterised by tremors and jerky movement [132]. The shoulder flexion feature requires a coordinated and sustained movement in a particular plane and as such it is logical that damage to such a region would be statistically significant.

Finally, the region of the putamen was found to be statistically significant for the elbow twist feature. The putamen is understood to have an important role in motor skills which include motor preparation, motor execution, and movement sequences. The

study by Crutcher et al [133] also found that neurons in the putamen of primates had activity which was strongly related to the direction of limb movements. This suggests that the putamen may have a primary role in limb control. The elbow pronation-supination feature requires coordinated rotation of the upper limb and dysfunction of this movement relates strongly with the current understanding of the putamen.

These statistically significant regions found with the gross motor outcome measures also correspond with the findings of prior LSM studies. One of the prevalent findings of these studies, as discussed in the literature review, was that subcortical lesions correlate more strongly with motor deficits than cortical lesions [119]. This result is supported in the present study with no statistically significant regions found in the cortex (despite this being a commonly affected region in the subject cohort as shown in Figure 4.10). Regions that were detected for gross motor outcomes were the subcortical regions of the CST and internal capsule. The CST is a well-known motor pathway and damage in this tract is known to have a major influence on motor dysfunction [134, 129]. Damage to the internal capsule has also been found to be a statistically significant region of upper extremity motor dysfunction as shown in several LSM studies [135, 99, 120]. The study by Frenkel-Toledo et al [98] also supports the finding between upper extremity dysfunction and damage to the putamen. Finally, damage to pallidum has been found to relate to balance problems in a prior LSM study [106] but as far as the author is aware, has not previously been connected to upper extremity motor function.

The findings detailed in the section align well with the findings of prior LSM studies. Damage to the CST, internal capsule, and putamen were found to be significantly related to gross motor dysfunction in the present study and these results are widely presented in the literature. Although these results do not provide any new scientific information, they do go some way to validating the usefulness of the clinical features. The pallidum and putamen are regions that were detected for the clinical features but not for the overall clinical score. This indicates the potential of the clinical features to provide a more fine-grained assessment of subcomponents of motor function. The pallidum has also not previously been identified as a region related to upper extremity motor function in prior LSM studies. This region was identified to be related to shoulder flexion in the current study indicating its potential involvement in this subcomponent of motor function. This novel insight indicates the potential of sensor-derived features to improve statistical findings in this domain

Clinical Features- Components of Fine Motor Function

The clinical features representing subcomponents of fine motor function are represented in Table 4.2 and Figures 4.14 and 4.15. Statistically significant regions were found which both align with existing novel and potentially offer a novel insight.

The CST was a statistically significant region for both the wrist flexion-extension and circumduction features. The CST is known to have a major role in fine motor function (as discussed in the Background section) so it follows that this region would be statistically significant.

The internal capsule and cortico-ponto-cerebellar tract were also found to be statically significant for the wrist flexion-extension task. The CST passes through the internal capsule and as such it is logical that this region would have an important role in both gross and fine motor function. The cortico-ponto-cerebellar tract is a more novel finding, and its functions are less well understood. A case study of isolated damage to this region in a study by Jang et al [136] found that the subject presented with resting and intentional tremor in both hands. This finding supports the results of the study that the cortico-ponto-cerebellar tract may have a critical role in hand/wrist motor function.

There are a limited number of LSM studies that have focused on upper extremity fine motor function which makes direct comparisons of results difficult. The CST was a region that was statistically significant for both fine motor outcome measures and this is reflected in the literature [97]. The internal capsule was also found to relate to poor upper extremity fine motor function and this has region has been demonstrated to result in poor recovery of hand function in a prior study [120]. Thus far, there have been no LSM studies that have reported statistically significant regions in the cortico-ponto-cerebellar tract related to motor function.

The features related to fine motor function in the present study found significance in the regions of the CST and internal capsule, findings which are widely published in the literature. The influence of the cortico-ponto-cerebellar tract on fine motor function is less widely published but is consistent with current knowledge about the function of this pathway. The influence of this region on fine motor function should be further investigated in the future. These results serve to both validate and show the potential of the clinical features to uncover new insights into the influence of certain brain regions on fine motor function.

4.13.2 Study Limitations

The early region-of-interest results provided by this novel motor outcome based LSM study are highly promising. The results indicate both the validity of the sensor-derived feature set and their potential to uncover more insights in future studies by providing fine-grained motor outcome measures. Despite this, there were a few limitations of the present study which should be improved upon in the future.

Three of the clinical features failed to attain statistically significant results. This may be due to the sample size implemented in the present study or may indicate an intrinsic fault with the features selected. One feature in particular, the shoulder twist rotation, did not show high usefulness as a measure of motor function (see Table 3.5) and failed to detect significance in the present study. As discussed in the prior chapter this movement is difficult to isolate and as a result this motor function is likely poorly captured. The two remaining features which failed to find significance in this study scored much higher in the test of usefulness in the previous chapter. The reason that these two features did not perform well in the present study is less obvious but is perhaps due to insufficient presence of damage to regions related to the dysfunction of these motor components.

As mentioned previously, one of the limitations of this study is the limited sample size (N=51) which is on the lower end of typical LSM studies. Although a larger cohort of subjects was initially recruited, many of these did not present with acute/ subacute lesions and as such had to be removed from the study.

One effect of the limited sample size of this study was that it was not possible to perform a VLSM study. This would have offered the potential to uncover more detailed information about the statically significant brain regions than was possible using the region-of-interest analysis in the present study.

Finally, there are the limitations associated with performing all analyses on the region- rather than voxel-level. This was a consequence of the aforementioned limited sample size which did not provide the necessary statistical power to compute insights at the voxel-level. One voxel-based method which could have been performed with a large sample size is the mass-univariate VLSM method. This approach would have enabled an investigation of the relationship between the features and specific brain locations (rather than pre-mapping brain regions) but does suffer from limitations including the "partial injury problem" and an assumption of independence between each voxel (as discussed in greater depth in the Background section). Another voxel-based technique which could have provided greater insight to the analyses are machine-learning based multivariate

models. These models offer the ability to utilise voxel-level information while minimising the limitations associated with the mass univariate approach.

CHAPTER

FIVE

Conclusions

The overarching goal of this thesis was to develop an automated system capable of providing a fine-grained assessment of motor function for application in LSM studies. The first study developed a wearable system composed of IMU and MMG sensors. This system provided outputs which combined the validity of the clinical rating scale with a series of fine-grained motor outcome measures. A second study conducted successfully implemented these outcome measures in a large LSM study.

The success of the developed wearable automated system was evaluated in terms of improving upon the major limitations of prior studies, the performance of the classifier, and the usefulness of the clinical feature set. The developed system addressed the drawbacks of previous automated studies by developing a system which could be used to test subjects at all motor function levels and could be used to quantify hand/wrist tasks. In addition, the system was tested on a much larger cohort of subjects than in the vast majority of prior automated studies.

The classification results achieved (in terms of accuracy and F-score) compared favourably to prior automated studies. The clinical features extracted from the orientation and myographic data all showed good correlation scores and are expected to offer a more fine-grained assessment of motor function than classified scores. The study proposes that a combination of classified score and clinical features would provide both the validity of the established clinical rating scores with the accuracy provided by the wearable sensor system.

A second major contribution provided by the developed automated system was the novel application of MMG sensors to quantify hand/wrist tasks. This region is a major contributor to upper extremity motor dysfunction post-stroke but has been ignored or

poorly quantified in most prior automated systems. Previous kinematic solutions have either performed poorly when quantifying this region (Kinect sensor) or have sizing and hygiene concerns (instrumented gloves). Recently myographic solutions have been proposed in the form of EMG sensors and these systems have achieved good results. This is the first study to implement MMG sensors, which have advantages over EMG including being more robust and reusable.

The early region-of-interest results provided by this novel motor outcome based LSM study are highly promising. Several of the features showed their validity for this application by the detection of significant regions that are consistent with current neuroscientific knowledge and the findings of prior LSM studies. This includes the relationship between gross motor function and damage to the subcortical region of the internal capsule and the CST. In addition, the clinical features evidenced sensitivity in the detection of regions related to subcomponents of motor function which broad clinical rating scales do not provide. The features for shoulder flexion and elbow twist detected statistically significant regions in the pallidum and putamen respectively, but these same regions were not detected by the clinical rating scale. Finally, there is some evidence that the fine-grained nature of the clinical features may lead to new insights not possible with clinical rating scales. For instance, the feature of wrist flexion-extension (relating to fine motor function) detected the significant region of the cortico-ponto-cerebellar tract. As far as the authour is aware, this finding has not been reported in a previous LSM study but is consistent with the symptoms reported by a case study documenting damage to this region.

CHAPTER

SIX

Future Work

In this chapter, the two major studies reported in this thesis will be discussed with respect to their limitations and how these should be addressed in any future work. The main limitations identified are the cohort size, improvements to the clinical feature set, and an expansion of the sensor system.

One limitation of both studies which could be rectified in the future is the cohort size. Although a relatively large pool of subjects was recruited (N=64 before exclusions), an even larger subject size would be expected to improve the analyses of both major studies presented in this paper. For the validation study of the novel automated system, an increase in sample size would be expected to solve several of the issues encountered with the classification model. This includes the bias issue whereby certain tasks of the clinical rating scale were found present with a large imbalance in class labels. A larger (and more balanced) training set would be expected to improve classification accuracy as well as the F-score achieved for each class. For the large-scale LSM study, recruiting further subjects to the study would enable a voxel- rather than region-level analysis to be performed. The region-of-interest analysis revealed a number of statistically significant regions when testing with the clinical score and clinical features. A larger sample size would enable a subsequent voxel-based analysis to be performed which may provide higher resolution anatomical information.

Another aspect of the studies which could be improved upon in the future is the clinical features utilised in both studies. The features were selected from a larger feature set based on three major criteria of transparency, usefulness, and generability. The criterion of usefulness was assessed by correlation with the clinical rating score, but the other two criteria were determined in a somewhat subject manner and this selection

criteria could certainly be improved in the future. Secondly, the total feature set, from which the clinical features were selected, is far from exhaustive and there are certainly other clinically relevant features which could be determined in the future.

A final limitation which could be improved in the future is the sensor system. The sensor system implemented provided good coverage of the upper arm with inertial sensing and flexion of the hand/wrist with myographic (MMG) sensing. One important movement this set up does not cover is wrist and finger extension. These movements are commonly affected post-stroke and also form a major component of several of the FMA-UE tasks. One change to capture this movement would be the addition of myographic (MMG) sensors positioned on the extrinsic extensors for the wrist and hand. Examples of muscles which could be instrumented in the future include the extensor carpi radialis longus and extensor digitorum.

CHAPTER

SEVEN

Clinical Implications

The work conducted in this thesis has outlined the development of an automated system and shown its potential for application in LSM mapping studies. One area in which the author believes this system (and other prior developed automated systems) are currently unsuitable is as a replacement for clinical assessment as part of normal post-stroke care. Although every effort made in the present study to ensure testing was as seamless as possible, the application of the system still presents an increased time, cognitive, and physical burden on clinician and subject alike, as compared to standard clinical testing. This is because these sensor systems require preparation time (such as fitting of sensors) and require the subjects to perform the tasks correctly in order to be able to collect useful sensor data. For this reason, it is considered that there is a low likelihood of widespread uptake by clinicians of such sensor systems for routine clinical evaluation.

The better application for the developed sensor system is in a field whereby more time may be afforded, and the high resolution provided by the fine-grained motor feature set may lead to useful insights. This includes the application in LSM studies as already demonstrated within the thesis. In addition, the developed automated system could be used to enrich studies of medicinal or rehabilitative interventions by improving the set of motor outcomes. The clinical features generated by the system offer a higher resolution measure of motor function than is possible with the standard clinical rating scores. If a subsequent study can prove, as expected, that this increased resolution translates to the ability to discriminate more levels of motor dysfunction, then this system could be applied as a highly sensitive measure of motor function. This could in turn enable intervention studies post-stroke to achieve statistical significance at smaller sample sizes or with only modest improvements in motor function.

Overall, the results presented in this thesis show significant promise in the developed automated system as a means to provide higher quality motor outcome metrics than currently possible using standard clinical rating scales. A novel system was developed which incorporated new findings in the application and feature extraction within MMG sensing, and a new measure of motor dysfunction which combines machine learning with clinically relevant parameters of motor function.

CHAPTER

EIGHT

Appendix A: Subject Demographics

Table 8.1: Clinical demographics for the cohort recruited for the wearable automated system and lesion-symptom mapping study

UID	Sex (M or F)	Age (years)	Affected side	FMA-UE score
2	F	80	RIGHT	40
4	F	68	LEFT	48
5	F	65	LEFT	58
6	M	43	LEFT	0
7	M	53	LEFT	15
8	M	60	RIGHT	30
9	F	85	RIGHT	55
10	M	64	RIGHT	35
11	F	69	LEFT	0
12	M	63	LEFT	40
13	F	78	RIGHT	47
14	F	70	RIGHT	32
15	F	83	LEFT	57
16	F	85	LEFT	34
17	F	95	LEFT	31
18	M	72	LEFT	8
19	M	58	LEFT	58
20	F	78	RIGHT	58
21	M	50	RIGHT	18
22	M	57	RIGHT	59
23	M	52	LEFT	12
24	M	60	LEFT	57
25	M	73	LEFT	60
26	M	44	RIGHT	9
27	M	40	LEFT	57
28	M	33	LEFT	11

UID	Sex (M or F)	Age (years)	Affected side	FMA-UE score
29	F	81	RIGHT	40
30	F	72	LEFT	20
31	F	75	LEFT	54
33	F	90	LEFT	57
34	M	75	LEFT	60
35	M	85	LEFT	54
36	F	42	LEFT	59
37	M	80	LEFT	58
39	M	73	LEFT	38
40	M	74	LEFT	50
41	F	63	LEFT	53
42	M	53	LEFT	56
43	F	80	RIGHT	59
44	M	53	RIGHT	18
45	F	79	RIGHT	11
46	M	54	LEFT	29
47	F	57	LEFT	30
48	M	48	LEFT	54
49	F	51	RIGHT	49
50	F	81	LEFT	2
51	F	82	LEFT	54
52	F	72	RIGHT	57
53	M	66	LEFT	54
54	F	47	RIGHT	59
55	M	54	LEFT	58
56	F	83	LEFT	45
57	F	77	LEFT	55
58	F	76	LEFT	53
59	M	64	LEFT	44
60	M	72	LEFT	51
61	M	60	LEFT	50
62	F	58	RIGHT	50
63	M	63	LEFT	58
64	M	66	LEFT	58
65	F	51	LEFT	43
66	F	75	RIGHT	55
67	M	66	RIGHT	54

Table 8.2: Clinical demographics and Imaging Specifications for the cohort recruited for the lesion-symptom mapping study. Stroke phase is determined at the time of motor outcome measure collection and based on periods specified in Table 2.2. Trials excluded from the study are indicated in the final column

UID	Stroke phase	Imaging modality	Excluded from the study?
2	ACUTE	CT	NO
4	SUBACUTE	MRI	NO
5	ACUTE	CT	NO
6	SUBACUTE	MRI	NO
7	SUBACUTE	CT	NO
8	ACUTE	MRI	NO
9	ACUTE	MRI	NO
10	ACUTE	MRI	NO
11	SUBACUTE	MRI	NO
12	ACUTE	MRI	NO
13	ACUTE	CT	NO
14	ACUTE	MRI	NO
15	ACUTE	MRI	NO
16	SUBACUTE	MRI	NO
17	ACUTE	MRI	NO
18	ACUTE	CT	NO
19	ACUTE	MRI	NO
20	SUBACUTE	MRI	NO
21	SUBACUTE	MRI	NO
22	ACUTE	MRI	NO
23	SUBACUTE	CT	NO
24	ACUTE	MRI	NO
25	ACUTE	MRI	NO
26	SUBACUTE	MRI	NO
27	ACUTE	MRI	NO
28	CHRONIC	MRI	YES

UID	Stroke phase	Imaging modality	Excluded from the study?
29	ACUTE	MRI	NO
30	SUBACUTE	CT	NO
31	SUBACUTE	CT	NO
33	CHRONIC	MRI	YES
34	ACUTE	MRI	NO
35	SUBACUTE	MRI	NO
36	ACUTE	MRI	NO
37	CHRONIC	CT	YES
39	CHRONIC	MRI	YES
40	ACUTE	MRI	NO
41	CHRONIC	MRI	YES
42	ACUTE	MRI	NO
43	CHRONIC	MRI	YES
44	ACUTE	MRI	NO
45	CHRONIC	CT	YES
46	SUBACUTE	MRI	NO
47	ACUTE	CT	NO
48	CHRONIC	MRI	YES
49	ACUTE	MRI	NO
50	ACUTE	MRI	NO
51	ACUTE	MRI	NO
52	CHRONIC	MRI	YES
53	ACUTE	MRI	NO
54	ACUTE	MRI	NO
55	CHRONIC	MRI	YES
56	ACUTE	CT	NO
57	CHRONIC	MRI	YES
58	ACUTE	MRI	NO
59	ACUTE	MRI	NO
60	ACUTE	MRI	YES
61	ACUTE	MRI	NO
62	CHRONIC	MRI	YES
63	ACUTE	MRI	NO
64	ACUTE	MRI	NO
65	ACUTE	MRI	NO
66	SUBACUTE	MRI	NO
67	ACUTE	MRI	NO

CHAPTER

NINE

Appendix B: Data Visualisation

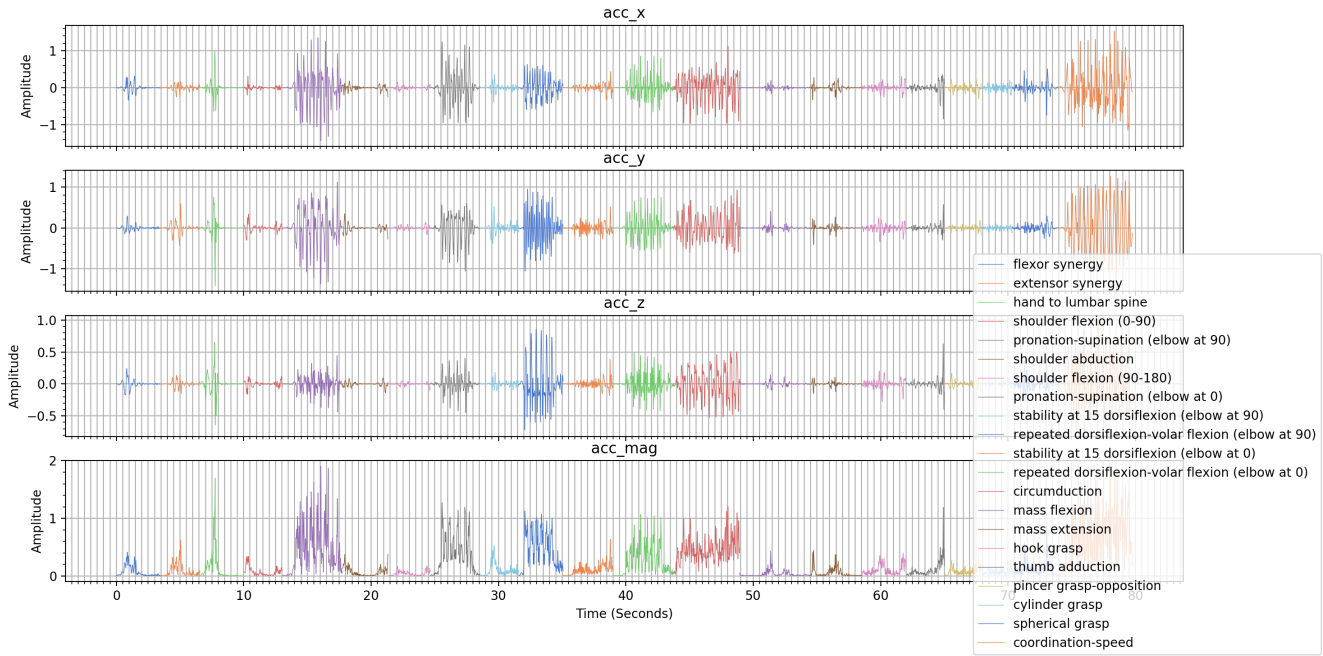


Figure 9.1: Data segmentation applied to accelerometer data recorded from the right lower arm of the subject during testing. Graphs are shown for each axis and magnitude of acceleration. Legend shows tasks of the FMA-UE clinical scale which have been segmented

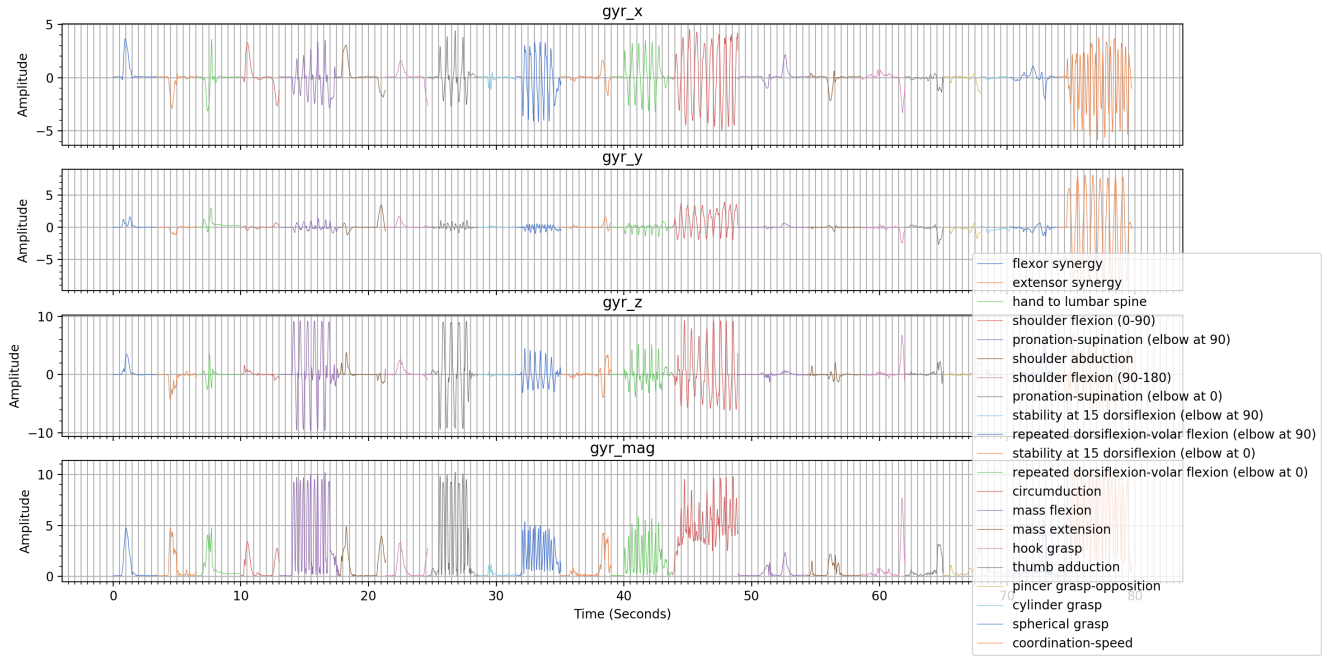


Figure 9.2: Data segmentation applied to gyroscope data recorded from the right lower arm of the subject during testing. Graphs are shown for each axis and magnitude of acceleration. Legend shows tasks of the FMA-UE clinical scale which have been segmented

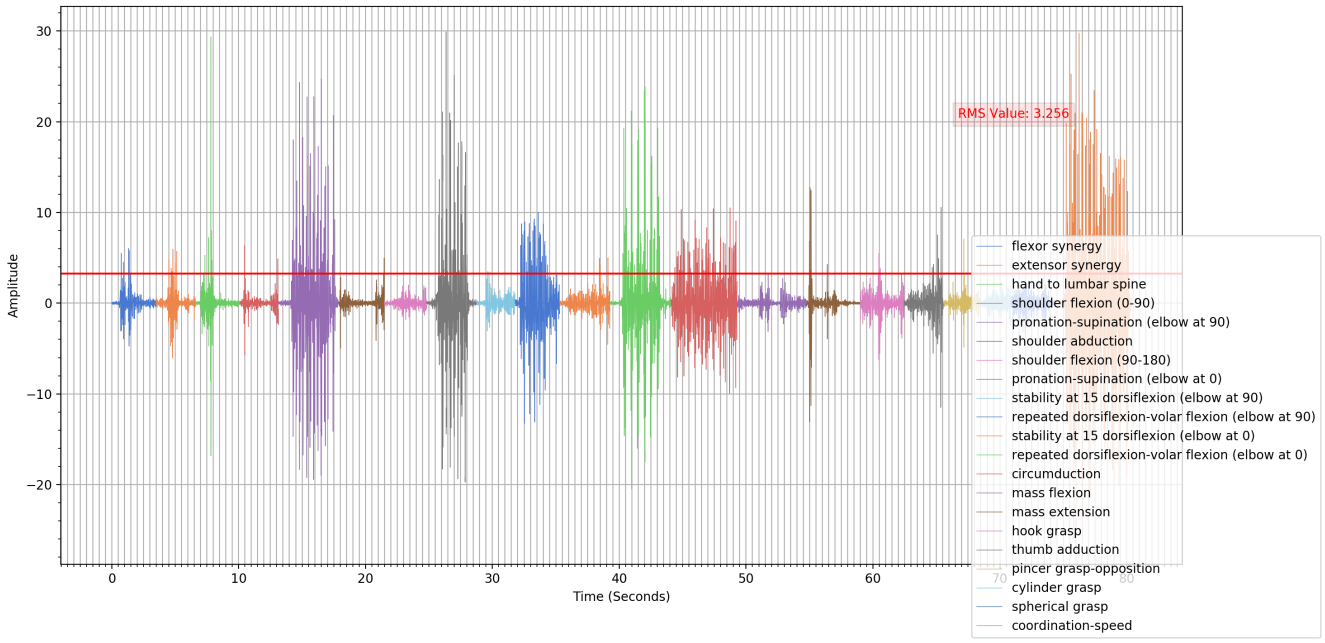


Figure 9.3: Data segmentation applied to mechanomyographic data recorded from the right lower arm of the subject during testing. Legend illustrates the tasks of the FMA-UE clinical scale which have been segmented

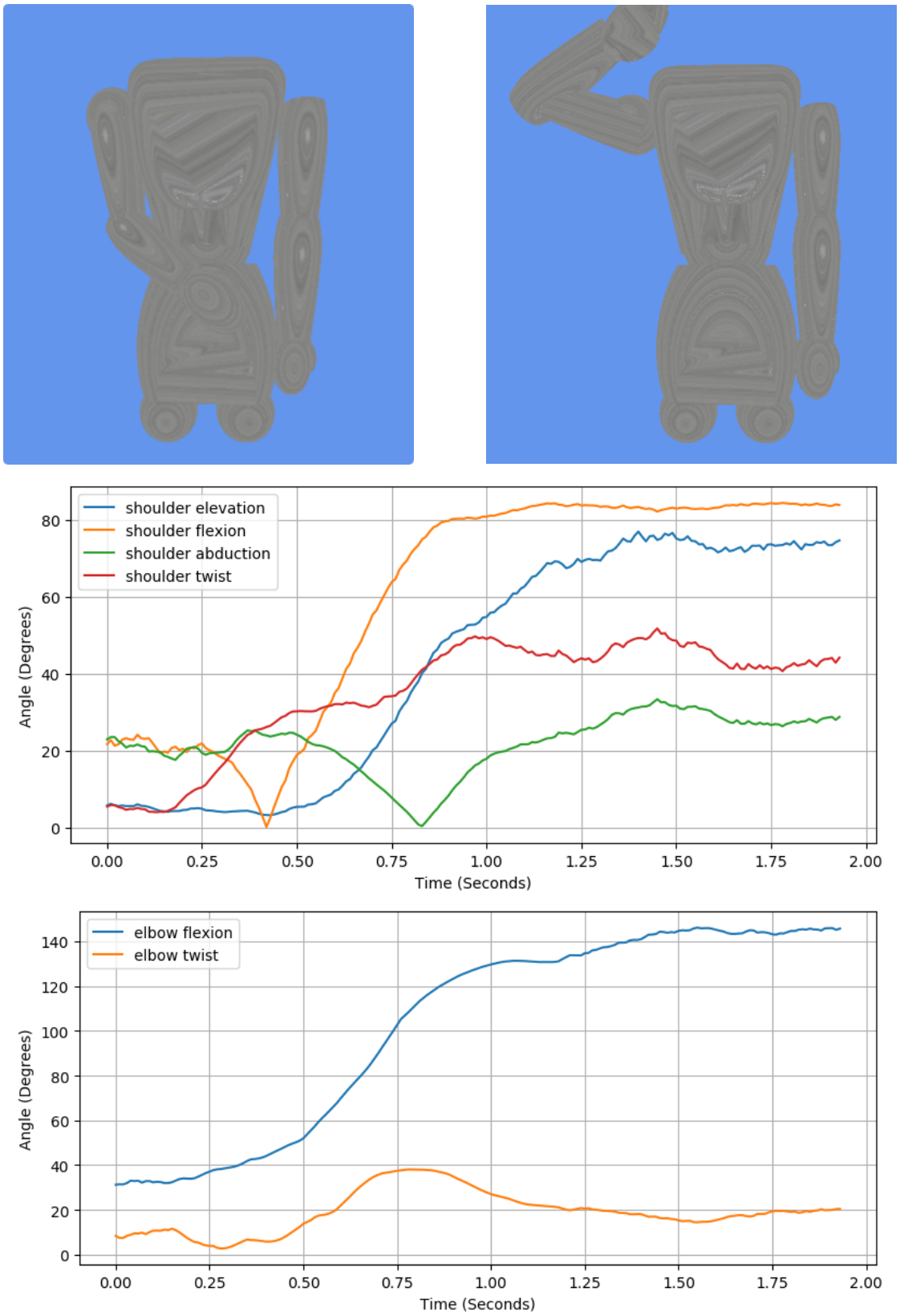


Figure 9.4: The simulated avatar and orientation readings of the shoulder and elbow during the "Flexor Synergy" task of the FMA-UE

CHAPTER

TEN

Appendix C: Confusion Matrices from the Wearable Automated System

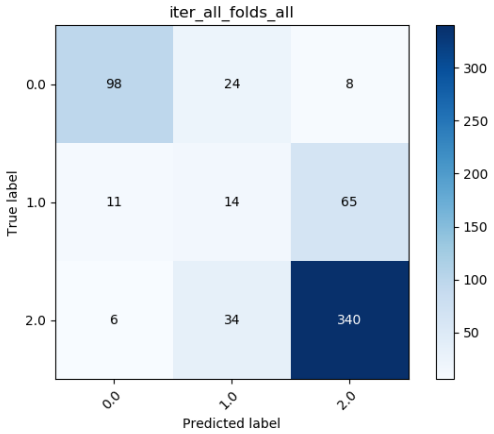


Figure 10.1: FMA-UE Confusion Matrix- "Flexor Synergy" Task

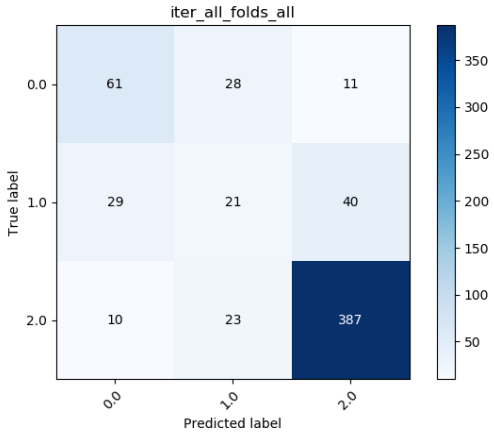


Figure 10.2: FMA-UE Confusion Matrix- "Extensor Synergy" Task

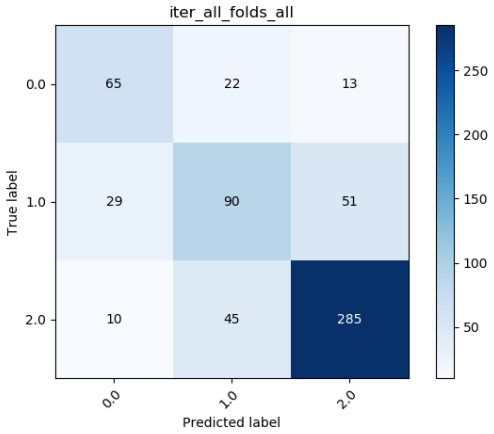


Figure 10.3: FMA-UE Confusion Matrix- "Hand to Lumbar Spine" Task

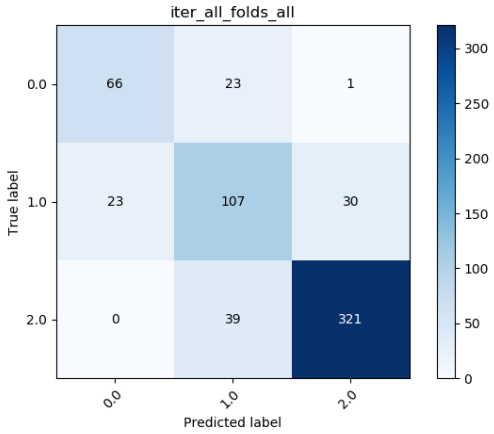


Figure 10.4: FMA-UE Confusion Matrix- "Shoulder Flexion (0-90 degrees)" Task

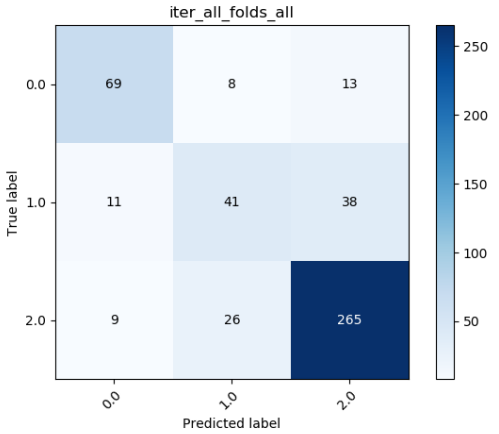


Figure 10.5: FMA-UE Confusion Matrix- "Pronation-Supination (elbow at 90 degrees)" Task

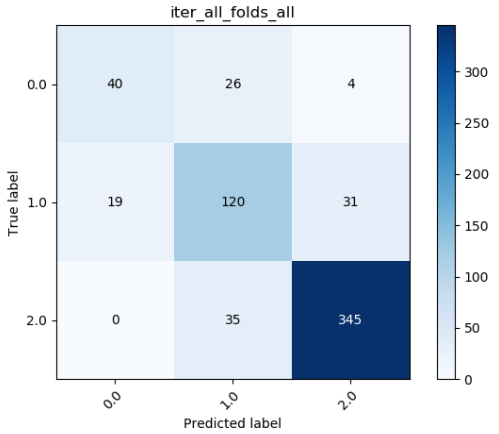


Figure 10.6: FMA-UE Confusion Matrix- "Shoulder Abduction" Task

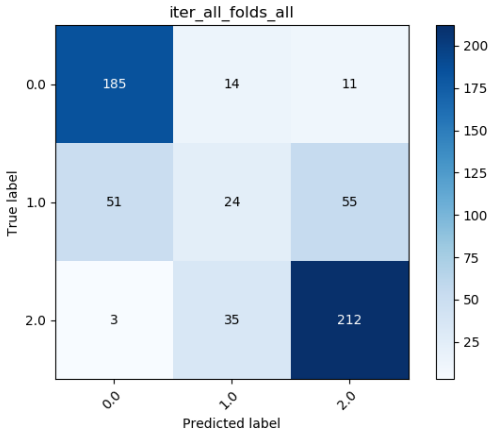


Figure 10.7: FMA-UE Confusion Matrix- "Shoulder Flexion (90-180 degrees)" Task

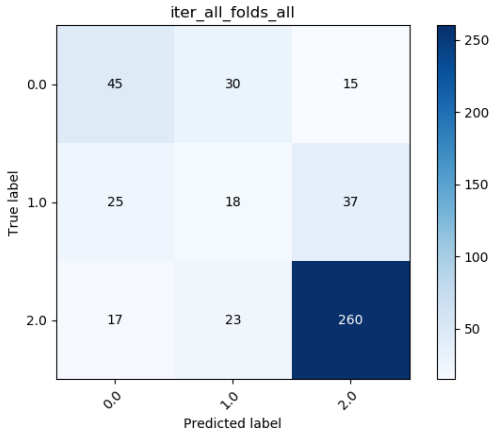


Figure 10.8: FMA-UE Confusion Matrix- "Pronation-Supination (elbow at 0 degrees)" Task

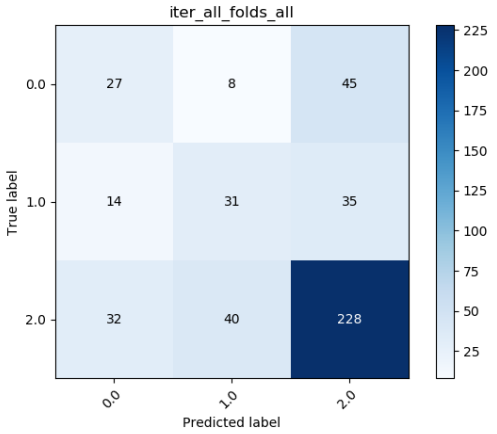


Figure 10.9: FMA-UE Confusion Matrix- "Stability at 15 Degrees Dorsiflexion (elbow at 90 degrees)" Task

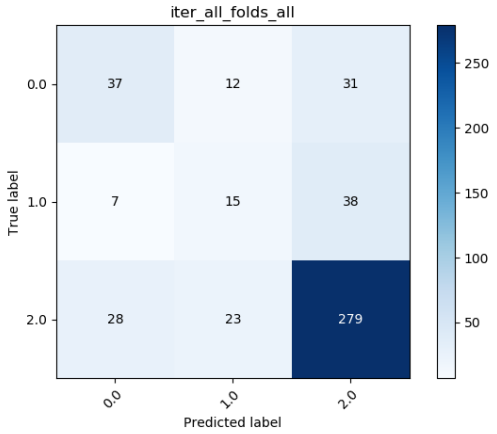


Figure 10.10: FMA-UE Confusion Matrix- "Repeated dorsi-volar flexion (elbow at 90 degrees)" Task

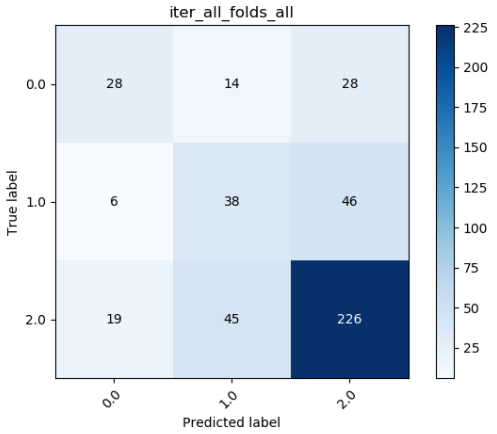


Figure 10.11: FMA-UE Confusion Matrix- "Stability at 15 degrees dorsiflexion (elbow at 0 degrees)" Task

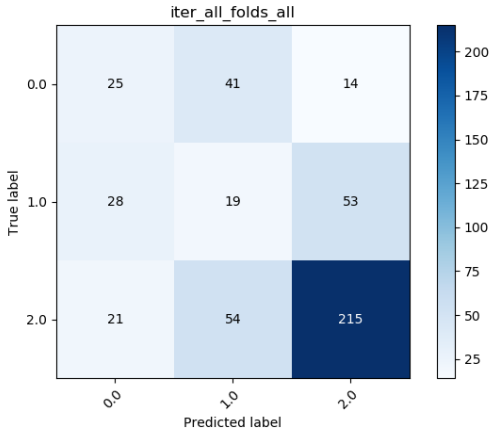


Figure 10.12: FMA-UE Confusion Matrix- "Repeated Dorsi-volar Flexion (elbow at 0 degrees)" Task

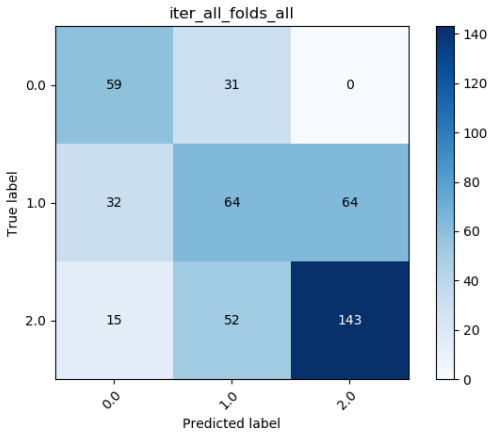


Figure 10.13: FMA-UE Confusion Matrix- "Circumduction" Task

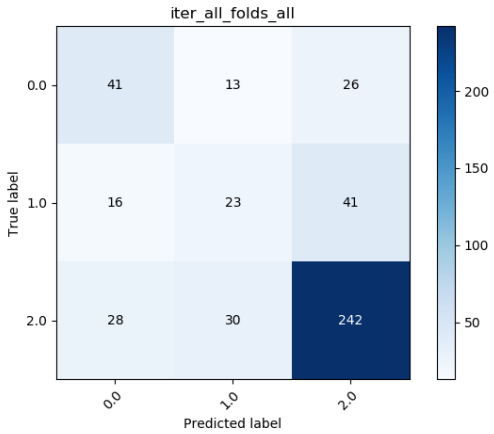


Figure 10.14: FMA-UE Confusion Matrix- "Thumb Adduction" Task

Bibliography

[1] "Types of Stroke," 2008.

[2] K. O. Lövblad, S. Altrichter, V. Mendes Pereira, M. Vargas, A. Marcos Gonzalez, S. Haller, and R. Sztajzel, "Imaging of acute stroke: CT and/or MRI," *Journal of Neuroradiology*, vol. 42, no. 1, pp. 55–64, 2015.

[3] "The Central Nervous System."

[4] P. M. Roos, "Studies on metals in motor neuron disease," Ph.D. dissertation, 2013.

[5] J. Konczak, D. Pierscianek, S. Hirsiger, U. Bultmann, B. Schoch, E. R. Gizewski, D. Timmann, M. Maschke, and M. Frings, "Recovery of upper limb function after cerebellar stroke: Lesion symptom mapping and arm kinematics," *Stroke*, vol. 41, no. 10, pp. 2191–2200, 2010.

[6] E. Bates, S. M. Wilson, A. P. Saygin, F. Dick, M. I. Sereno, R. T. Knight, and N. F. Dronkers, "Voxel-based lesion–symptom mapping," *Nature Neuroscience*, pp. 11–12, 2003.

[7] D. Tomaszewski, J. Rapiński, and R. Pelc-Mieczkowska, "Concept of AHRS Algorithm Designed for Platform Independent Imu Attitude Alignment," *Reports on Geodesy and Geoinformatics*, vol. 104, no. 1, pp. 33–47, 2018.

[8] A. Chakure, "Random Forest Classification," 2019.

[9] Z. Zhang, "Boosting Algorithms Explained," 2019.

[10] "Features," 2020.

[11] S. Patel, R. Hughes, T. Hester, J. Stein, M. Akay, J. Dy, and P. Bonato, "Tracking Motor Recovery in Stroke Survivors Undergoing Rehabilitation Using Wearable Technology," in *IEEE Engineering in Medicine and Biology Society*, 2010, pp. 6858–6861.

[12] V. F. Bento, V. T. Cruz, D. D. Ribeiro, and J. P. S. Cunha, "Towards a movement quantification system capable of automatic evaluation of upper limb motor function after neurological injury," in *IEEE Engineering in Medicine and Biology Society*, 2011, pp. 5456–5460.

[13] E. V. Olesh, S. Yakovenko, and V. Gritsenko, "Automated assessment of upper extremity movement impairment due to stroke," *PLoS ONE*, vol. 9, no. 8, 2014.

[14] C. Bosecker, L. Dipietro, B. Volpe, and H. I. Krebs, "Kinematic Robot-Based Evaluation Scales and Clinical Counterparts to Measure Upper Limb Motor Performance in Patients with Chronic Stroke," *Neurorehabilitation and Neural Repair*, vol. 24, no. 1, pp. 62–69, 2010.

[15] Y. Li, X. Zhang, Y. Gong, Y. Cheng, X. Gao, and X. Chen, "Motor Function Evaluation of Hemiplegic Upper-Extremities Using Data Fusion from Wearable Inertial and Surface EMG Sensors," *Sensors*, vol. 17, no. 3, p. 582, 2017.

[16] 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.

[17] S. Del Din, S. Patel, C. Cobelli, and P. Bonato, "Estimating fugl-meyer clinical scores in stroke survivors using wearable sensors," in *IEEE Engineering in Medicine and Biology Society*, 2011, pp. 5839–5842.

[18] J. Basmajian, *Biofeedback: Principles and Practice for Clinicians*, 3rd ed. Lippincott Williams and Wilkins, 1989.

[19] W. S. Kim, S. Cho, D. Baek, H. Bang, and N. J. Paik, "Upper extremity functional evaluation by Fugl-Meyer assessment scoring using depth-sensing camera in hemiplegic stroke patients," *PLoS ONE*, vol. 11, no. 7, pp. 1–13, 2016.

[20] S. Lee, Y. S. Lee, and J. Kim, "Automated Evaluation of Upper-Limb Motor Function Impairment Using Fugl-Meyer Assessment," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 1, pp. 125–134, 2018.

[21] B. de Haan, P. Clas, H. Juenger, M. Wilke, and H.-O. Karnath, "Fast semi-automated lesion demarcation in stroke." *NeuroImage: Clinical*, vol. 9, pp. 69–74, 2015.

[22] "State of the nation," Stroke Association, Tech. Rep., 2018.

[23] L. M. Brass, "STROKE," in *Yale University School of Medicine Heart Book*, 1992, pp. 215–233.

[24] P. Raghavan, "Upper Limb Motor Impairment Post Stroke," *Physical Medicine and Rehabilitation Clinics of North America*, vol. 26, no. 4, pp. 599–610, 2016.

[25] J. Chae, G. Yang, B. K. Park, and I. Labatia, "Delay in initiation and termination of muscle contraction, motor impairment, and physical disability in upper limb hemiparesis," *Muscle and Nerve*, vol. 25, no. 4, pp. 568–575, 2002.

[26] C. G. Canning, L. Ada, and N. O'Dwyer, "Slowness to develop force contributes to weakness after stroke," *Archives of Physical Medicine and Rehabilitation*, vol. 80, no. 1, pp. 66–70, 1999.

[27] N. Bolognini, C. Russo, and D. J. Edwards, "The sensory side of post-stroke motor rehabilitation," *Restorative Neurology and Neuroscience*, vol. 34, no. 4, pp. 571–586, 2016.

[28] C. L. Watkins, M. J. Leathley, J. M. Gregson, A. P. Moore, T. L. Smith, and A. K. Sharma, "Prevalence of spasticity post stroke," *Clinical Rehabilitation*, vol. 16, no. 5, pp. 515–522, 2002.

[29] C. L. Kuo and G. C. Hu, "Post-stroke Spasticity: A Review of Epidemiology, Pathophysiology, and Treatments," *International Journal of Gerontology*, vol. 12, no. 4, pp. 280–284, 2018.

[30] R. Beer, M. Ellis, B. Holubar, and J. Dewald, "IMPACT OF GRAVITY LOADING ON POST-STROKE REACHING AND ITS RELATIONSHIP TO WEAKNESS," *Muscle and Nerve*, vol. 36, no. 2, pp. 242–250, 2007.

[31] M. Alt Murphy, C. Resteghini, P. Feys, and I. Lamers, "An overview of systematic reviews on upper extremity outcome measures after stroke." *BMC Neurology*, vol. 15, no. 29, 2015.

[32] K. Salter, N. Campbell, M. Richardson, and S. Mehta, "Outcome measures in stroke rehabilitation," in *Handbook of Clinical Neurology*, 2013.

[33] A. Fugl-Meyer, L. Jääskö, I. Leyman, S. Olsson, and S. Steglind, "The post-stroke hemiplegic patient. 1. a method for evaluation of physical performance," *Scandinavian Journal of Rehabilitation Medicine*, vol. 7, no. 1, pp. 13–31, 1975.

[34] D. Birenbaum, L. W. Bancroft, and G. J. Felsberg, "Imaging in acute stroke," *Western Journal of Emergency Medicine*, vol. 12, no. 1, pp. 67–76, 2011.

[35] C. Sperber and H. O. Karnath, "On the validity of lesion-behaviour mapping methods," *Neuropsychologia*, vol. 115, no. August 2017, pp. 17–24, 2018.

[36] L. M. Allen, A. N. Hasso, J. Handwerker, and H. Farid, "Sequence-specific MR imaging findings that are useful in dating ischemic stroke," *Radiographics*, vol. 32, no. 5, pp. 1285–1297, 2012.

[37] "Anatomy of the Brain."

[38] K. B. Lee and S. H. Lim, "Prognosis and Recovery of Motor Function with Lesion–Symptom Mapping in Patients with Stroke," *Brain & Neurorehabilitation*, vol. 10, no. 1, 2017.

[39] A. Sterr, P. J. Dean, A. J. Szameitat, A. B. Conforto, and S. Shen, "Corticospinal tract integrity and lesion volume play different roles in chronic hemiparesis and its improvement through motor practice," *Neurorehabilitation and Neural Repair*, vol. 28, no. 4, pp. 335–343, 2014.

[40] R. Lindenberg, V. Renga, L. L. Zhu, F. Betzler, D. Alsop, and G. Schlaug, "Structural integrity of corticospinal motor fibers predicts motor impairment in chronic stroke," *Neurology*, vol. 74, no. 4, pp. 280–287, 2010.

[41] L. L. Zhu, R. Lindenberg, M. P. Alexander, and G. Schlaug, "Lesion load of the corticospinal tract predicts motor impairment in chronic stroke," *Stroke*, vol. 41, no. 5, pp. 910–915, 2010.

[42] E. A. Stamatakis and L. K. Tyler, "Identifying lesions on structural brain images - Validation of the method and application to neuropsychological patients," *Brain and Language*, vol. 94, no. 2, pp. 167–177, 2005.

[43] A. C. Evans, A. L. Janke, D. L. Collins, and S. Baillet, "Brain templates and atlases," *NeuroImage*, vol. 62, no. 2, pp. 911–922, 2012.

[44] M. Sarter, G. G. Berntson, and J. T. Cacioppo, "Brain Imaging and Cognitive Neuroscience: Toward Strong Inference in Attributing Function to Structure," *American Psychologist*, vol. 51, no. 1, pp. 13–21, 1996.

[45] S. M. Wilson, "Lesion-symptom mapping in the study of spoken language understanding," *Language, Congnition and Neuroscience*, vol. 32, no. 7, pp. 891–899, 2017.

[46] C. Rorden, J. Fridriksson, and H.-O. Karnath, "An evaluation of traditional and novel tools for lesion behavior mapping," *Neuroimage*, vol. 44, no. 4, pp. 1355–1362, 2009.

[47] D. Ao, R. Sun, and R. Song, "Comparison of complexity of EMG signals between a normal subject and a patient after stroke - A case study," in *IEEE Engineering in Medicine and Biology Society*, no. 4, 2013, pp. 4965–4968.

[48] A. Péter, E. Andersson, A. Hegyi, T. Finni, O. Tarassova, N. Cronin, H. Grundström, and A. Arndt, "Comparing Surface and Fine-Wire Electromyography Activity of Lower Leg Muscles at Different Walking Speeds," *Frontiers in Physiology*, vol. 10, no. October, pp. 1–12, 2019.

[49] G. Oster and J. Jaffe, "Low frequency sounds from sustained contraction of human skeletal muscle," *Biophysical Journal*, vol. 30, no. 1, pp. 119–127, 1980.

[50] M. O. Ibitoye, N. A. Hamzaid, J. M. Zuniga, and A. K. Abdul Wahab, "Mechanomyography and muscle function assessment: A review of current state and prospects," *Clinical Biomechanics*, vol. 29, no. 6, pp. 691–704, 2014.

[51] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Journal of Fluids Engineering, Transactions of the ASME*, vol. 82, no. 1, pp. 35–45, 1960.

[52] P. Bernal-Polo and H. Martínez-Barberá, "Kalman filtering for attitude estimation with quaternions and concepts from manifold theory," *Sensors (Switzerland)*, vol. 19, no. 1, 2019.

[53] X. Yun, E. R. Bachmann, R. B. Mcghee, and M. J. Zyda, "An Extended Kalman Filter for Quaternion-Based Orientation Estimation Using MARG Sensors," in *International Conference on Intelligent Robotics and Systems*, 2001, pp. 1–9.

[54] J. K. Uhlmann and S. J. Julier, "A New Extension of the Kalman Filter to Nonlinear Systems," *Signal processing, sensor fusion, and target recognition VI*, vol. 3068, pp. 182–194, 1997.

[55] S. O. H. Madgwick, A. J. L. Harrison, and R. Vaidyanathan, "Estimation of IMU and MARG orientation using a gradient descent algorithm," in *IEEE International Conference on Rehabilitation Robotics*, 2011, pp. 1–7.

[56] I. H. Witten, E. Frank, and M. a. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*, 2011.

[57] A. Lucas, J. Hermiz, J. Labuzetta, Y. Arabadzhi, N. Karanjia, and V. Gilja, "Use of Accelerometry for Long Term Monitoring of Stroke Patients," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 7, no. January, pp. 1–10, 2019.

[58] X. Liu, S. Rajan, N. Ramasarma, P. Bonato, and S. I. Lee, "The Use of a Finger-Worn Accelerometer for Monitoring of Hand Use in Ambulatory Settings," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 2, pp. 599–606, 2019.

[59] B. Knorr, R. Hughes, D. Sherrill, J. Stein, M. Akay, and P. Bonato, "Quantitative measures of functional upper limb movement in persons after stroke," in *IEEE/EMBS Conference on Neural Engineering*, vol. 2005, 2005, pp. 252–255.

[60] T. Hester, R. Hughes, D. Sherrill, B. Knorr, M. Akay, J. Stein, and P. Bonato, "Using wearable sensors to measure motor abilities following stroke," in *IEEE International Workshop on Wearable and Implantable Body Sensor Networks*, 2006, pp. 4–8.

[61] G. Wang, Z. Zhang, and C. Ayala, "Costs of Hospitalization for Stroke Patients Aged 18-64 Years in the United States," *Journal of Stroke and Cerebrovascular Disease*, vol. 23, no. 5, pp. 861–868, 2014.

[62] A. Parnandi, E. Wade, and M. Matari, "Motor Function Assessment Using Wearable Inertial Sensors," in *IEEE Engineering in Medicine and Biology Society*, 2010, pp. 86–89.

[63] M. Zhang, B. Lange, C.-y. Chang, A. A. Sawchuk, and A. A. Rizzo, "Beyond the Standard Clinical Rating Scales : Fine-Grained Assessment of Post-Stroke Motor Functionality Using Wearable Inertial Sensors," in *IEEE Engineering in Medicine and Biology Society*, 2012, pp. 6111–6115.

[64] S. Huang, C. Luo, S. Ye, F. Liu, B. Xie, C. Wang, L. Yang, Z. Huang, and J. Wu, "Motor impairment evaluation for upper limb in stroke patients on the basis of a microsensor," *International Journal of Rehabilitation Research*, vol. 35, no. 2, pp. 161–169, 2012.

[65] V. T. Cruz, V. F. Bento, D. D. Ribeiro, I. Araújo, C. A. Branco, and P. Coutinho, "A novel system for automatic classification of upper limb motor function after stroke: an exploratory study." *Medical Engineering & Physics*, vol. 36, no. 12, pp. 1704–10, 2014.

[66] A. Scano, A. Chiavenna, M. Malosio, L. Molinari Tosatti, and F. Molteni, "Kinect V2 implementation and testing of the reaching performance scale for motor evaluation of patients with neurological impairment," *Medical Engineering & Physics*, vol. 56, pp. 54–58, 2018.

[67] N. J. Seo, V. Crocher, E. Spaho, C. R. Ewert, M. F. Fathi, P. Hur, S. A. Lum, E. M. Humanitzki, A. L. Kelly, V. Ramakrishnan, and M. L. Woodbury, "Capturing upper limb gross motor categories using the kinect® sensor," *American Journal of Occupational Therapy*, vol. 73, no. 4, pp. 1–10, 2019.

[68] P. Otten, J. Kim, and S. H. Son, "A framework to automate assessment of upper-limb motor function impairment: A feasibility study," *Sensors*, vol. 15, no. 8, pp. 20 097–20 114, 2015.

[69] R. Julianjatsono, R. Ferdiana, and R. Hartanto, "High-resolution automated Fugl-Meyer Assessment using sensor data and regression model," in *IEEE International Conference on Science and Technology - Computer*, 2017, pp. 28–32.

[70] Y. Chongyang, W. Zixi, and J. Linhong, "A research on the feasibility of automatic Fugl-Meyer assessment for upper limb based on rehabilitation devices," in *International Convention on Rehabilitation Engineering & Assistive Technology*, 2008, pp. 126–129.

[71] S. Balasubramanian, R. Wei, R. Herman, and J. He, "Robot-measured performance metrics in stroke rehabilitation," in *IEEE International Conference on Complex Medical Engineering*, 2009, pp. 0–5.

[72] S. Subramaniam, R. Varghese, and T. Bhatt, "Influence of Chronic Stroke on Functional Arm Reaching: Quantifying Deficits in the Ipsilesional Upper Extremity," *Rehabilitation Research and Practice*, vol. 2019, 2019.

[73] W. Tang, X. Zhang, X. Tang, S. Cao, X. Gao, and X. Chen, "Surface Electromyographic Examination of Poststroke Neuromuscular Changes in Proximal and Distal Muscles Using Clustering Index Analysis," *Frontiers in Neurology*, vol. 8, p. 731, 2018.

[74] X. Zhang, Z. Wei, X. Ren, X. Gao, X. Chen, and P. Zhou, "Complex Neuromuscular Changes Post-Stroke Revealed by Clustering Index Analysis of Surface Electromyogram," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 11, pp. 2105–2112, 2017.

[75] H. Uesugi, M. Sonoo, E. Stålberg, K. Matsumoto, M. Higashihara, H. Murashima, Y. Ugawa, Y. Nagashima, T. Shimizu, H. Saito, and I. Kanazawa, ""Clustering Index method": A new technique for differentiation between neurogenic and myopathic changes using surface EMG," *Clinical Neurophysiology*, vol. 122, no. 5, pp. 1032–1041, 2011.

[76] X. L. Hu, K. Y. Tong, and L. Li, "The mechanomyography of persons after stroke during isometric voluntary contractions." *Journal of electromyography and kinesiology*, vol. 17, no. 4, pp. 473–83, 2007.

[77] B. Kim and C. Winstein, "Can Neurological Biomarkers of Brain Impairment Be Used to Predict Poststroke Motor Recovery? A Systematic Review," *Neurorehabilitation and Neural Repair*, vol. 31, no. 1, pp. 3–24, 2017.

[78] E. Repnik, U. Puh, N. Goljar, M. Munih, and M. Mihelj, "Using inertial measurement units and electromyography to quantify movement during action research arm test execution," *Sensors (Switzerland)*, vol. 18, no. 9, pp. 1–23, 2018.

[79] H. T. Jung, H. Kim, M. Y. Oh, T. Ryu, and Y. Kim, "Learning classifier to evaluate movement quality in unassisted pick-and-place exercises for post-stroke patients: A preliminary study," *IEEE Engineering in Medicine and Biology Society*, pp. 2490–2493, 2017.

[80] A. Scano, M. Caimmi, A. Chiavenna, M. Malosio, and L. M. Tosatti, "Kinect One-based biomechanical assessment of upper-limb performance compared to clinical scales in post-stroke patients," in *IEEE Engineering in Medicine and Biology Society*, 2015, pp. 5720–5723.

[81] K. Leuenberger, R. Gonzenbach, S. Wachter, A. Luft, and R. Gassert, "A method to qualitatively assess arm use in stroke survivors in the home environment," *Medical and Biological Engineering and Computing*, vol. 55, no. 1, pp. 141–150, 2017.

[82] M. Mace, S. Guy, A. Hussain, E. Diane Playford, N. Ward, S. Balasubramanian, and E. Burdet, "Validity of a sensor-based table-top platform to measure upper limb function," in *IEEE International Conference on Rehabilitation Robotics*. IEEE, 2017, pp. 652–657.

[83] B. Dehbandi, A. Barachant, D. Harary, J. D. Long, K. Z. Tsagaris, S. J. Bumanlag, V. He, and D. Putrino, "Using Data from the Microsoft Kinect 2 to Quantify Upper Limb Behavior: A Feasibility Study," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 5, pp. 1386–1392, 2017.

[84] G. V. Dijck, J. V. Vaerenbergh, and M. M. V. Hulle, "Posterior Probability Profiles for the Automated Assessment of the Recovery of Stroke Patients," in *Artificial Intelligence in Medicine*, 2007, pp. 347–353.

[85] N. Hesam-Shariati, T. Trinh, A. G. Thompson-Butel, C. T. Shiner, S. J. Redmond, and P. A. McNulty, "Improved Kinematics and Motor Control in a Longitudinal Study of a Complex Therapy Movement in Chronic Stroke," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 4, pp. 682–691, 2019.

[86] L. Jaasko, I. Leyman, S. Olsson, and S. Steglind, "FMA-UE Protocol," Tech. Rep., 2010.

[87] B. Rohrer, S. Fasoli, H. I. Krebs, R. Hughes, B. Volpe, W. R. Frontera, J. Stein, and N. Hogan, "Movement smoothness changes during stroke recovery," *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, vol. 22, no. 18, pp. 8297–8304, 2002.

[88] A. Posatskiy and T. Chau, "Design and evaluation of a novel microphone-based mechanomyography sensor with cylindrical and conical acoustic chambers," *Medical Engineering & Physics*, vol. 34, no. 8, pp. 1184–1190, 2012.

[89] FormLabs, "Materials Data Sheet," Tech. Rep., 2018.

[90] M. Gor-Garcia-Fogeda, F. Molina-rueda, A. Cuesta-Gomez, I. M. Alguacil-diego, and J. C. Miangolarra-page, "Scales to Assess Gross Motor Function in Stroke Patients : A Systematic Review," *Archives of Physical Medicine and Rehabilitation*, vol. 95, pp. 1174–83, 2014.

[91] M. O. Ibitoye, N. A. Hamzaid, J. M. Zuniga, N. Hasnan, and A. K. A. Wahab, "Mechanomyographic parameter extraction methods: An appraisal for clinical applications," *Sensors*, vol. 14, no. 12, pp. 22 940–22 970, 2014.

[92] C. Orizio, "Muscle sound: bases for the introduction of a mechanomyographic signal in muscle studies." *Critical reviews in biomedical engineering*, vol. 21, no. 3, pp. 201–243, 1993.

[93] A. Phinyomark, P. Phukpattaranont, and C. Limsakul, "Feature reduction and selection for EMG signal classification," *Expert Systems with Applications*, vol. 39, no. 8, pp. 7420–7431, 2012.

[94] M. Ojala and G. C. Garriga, "Permutation tests for studying classifier performance," *Journal of Machine Learning Research*, vol. 11, pp. 1833–1863, 2010.

[95] 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. 232–40, sep 2002.

[96] T. S. T. Fu, C. Y. Wu, K. C. Lin, C. J. Hsieh, J. S. Liu, T. N. Wang, and P. Ou-Yang, "Psychometric comparison of the shortened Fugl-Meyer Assessment and the streamlined Wolf Motor Function Test in stroke rehabilitation," *Clinical Rehabilitation*, vol. 26, no. 11, pp. 1043–1047, 2012.

[97] J. Plantin, G. V. Pennati, P. Roca, J. C. Baron, E. Laurencikas, K. Weber, A. K. Godbolt, J. Borg, and P. G. Lindberg, "Quantitative assessment of hand spasticity after stroke: Imaging correlates and impact on motor recovery," *Frontiers in Neurology*, vol. 10, no. JUL, pp. 1–11, 2019.

[98] S. Frenkel-Toledo, G. Fridberg, S. Ofir, G. Bartur, J. Lowenthal-Raz, O. Granot, S. Handelzalts, and N. Soroker, "Lesion location impact on functional recovery of the hemiparetic upper limb," *PLoS ONE*, vol. 14, no. 7, pp. 1–28, 2019.

[99] K. B. Lee, J. S. Kim, B. Y. Hong, and S. H. Lim, "Clinical recovery from stroke lesions and related outcomes," *Journal of Clinical Neuroscience*, vol. 37, pp. 79–82, 2017.

[100] J. A. Fiez, H. Damasio, and T. J. Grabowski, "Lesion segmentation and manual warping to a reference brain: Intra- and interobserver reliability," *Human Brain Mapping*, vol. 9, no. 4, pp. 192–211, 2000.

[101] S. Baillieul, C. Elsworth-Edelsten, A. Saj, and G. Allali, "Neural substrates of reduced walking activity after supratentorial stroke: A voxel-based lesion symptom mapping study," *Human Movement Science*, vol. 67, no. May, p. 102517, 2019.

[102] J. Wilmskoetter, L. Bonilha, B. Martin-Harris, J. J. Elm, J. Horn, and H. S. Bonilha, "Mapping acute lesion locations to physiological swallow impairments after stroke," *NeuroImage: Clinical*, vol. 22, 2019.

[103] H. I. Moon, H. J. Lee, and S. Y. Yoon, "Lesion location associated with balance recovery and gait velocity change after rehabilitation in stroke patients," *Neuroradiology*, vol. 59, no. 6, pp. 609–618, 2017.

[104] K. M. Oostra, A. Van Bladel, A. C. L. Vanhoonacker, and G. Vingerhoets, "Damage to Fronto-Parietal Networks Impairs Motor Imagery Ability after Stroke: A Voxel-Based Lesion Symptom Mapping Study." *Frontiers in Behavioral Neuroscience*, vol. 10, no. February, p. 5, 2016.

[105] L. M. Skipper-Kallal, E. H. Lacey, S. Xing, and P. E. Turkeltaub, "Functional activation independently contributes to naming ability and relates to lesion site in post-stroke aphasia," *Human Brain Mapping*, vol. 38, no. 4, pp. 2051–2066, 2017.

[106] A. M. Reynolds, D. M. Peters, J. M. Vendemia, L. P. Smith, R. C. Sweet, G. C. Baylis, D. Krotish, and S. L. Fritz, "Neuronal injury in the motor cortex after chronic stroke and lower limb motor impairment: A voxelbased lesion symptom mapping study," *Neural Regeneration Research*, vol. 9, no. 7, pp. 766–772, 2014.

[107] S. Vossel, P. H. Weiss, P. Eschenbeck, and G. R. Fink, "Anosognosia, neglect, extinction and lesion site predict impairment of daily living after right-hemispheric stroke," *Cortex*, vol. 49, no. 7, pp. 1782–1789, 2013.

[108] L. D. Alexander, S. E. Black, F. Gao, G. Szilagyi, C. J. Danells, and W. E. Mcilroy, "Correlating lesion size and location to deficits after ischemic stroke: the influence of accounting for altered peri-necrotic tissue and incidental silent infarcts," *Behaviour and Brain Functions*, vol. 6, no. 6, pp. 1–10, 2010.

[109] K. L. Ito, H. Kim, and S. L. Liew, "A comparison of automated lesion segmentation approaches for chronic stroke T1-weighted MRI data," *Human Brain Mapping*, vol. 40, no. 16, pp. 4669–4685, 2019.

[110] S. L. Liew, J. M. Anglin, N. W. Banks, M. Sondag, K. L. Ito, H. Kim, J. Chan, J. Ito, C. Jung, N. Khoshab, S. Lefebvre, W. Nakamura, D. Saldana, A. Schmiesing, C. Tran, D. Vo, T. Ard, P. Heydari, B. Kim, L. Aziz-Zadeh, S. C. Cramer, J. Liu, S. Soekadar, J. E. Nordvik, L. T. Westlye, J. Wang, C. Winstein, C. Yu, L. Ai, B. Koo, R. C. Craddock, M. Milham, M. Lakich, A. Pienta, and A. Stroud, "A large, open source dataset of stroke anatomical brain images and manual lesion segmentations," *Scientific Data*, vol. 5, pp. 1–11, 2018.

[111] M. S. Babu and V. Vijayalakshmi, "A review on acute/sub-acute ischemic stroke lesion segmentation and registration challenges," *Multimedia Tools and Applications*, vol. 78, no. 2, pp. 2481–2506, 2019.

[112] M. Wilke, B. de Haan, H. Juenger, H.-o. O. Karnath, B. D. Haan, H. Juenger, H.-o. O. Karnath, B. de Haan, H. Juenger, H.-o. O. Karnath, B. D. Haan, H. Juenger, and H.-o. O. Karnath, "Manual, semi-automated, and automated delineation of chronic brain lesions: A comparison of methods," *NeuroImage*, vol. 56, no. 4, pp. 2038–2046, 2011.

[113] M. Joliot, G. Jobard, M. Naveau, N. Delcroix, L. Petit, L. Zago, F. Crivello, E. Mellet, B. Mazoyer, and N. Tzourio-Mazoyer, "AICHA: An atlas of intrinsic connectivity of homotopic areas," *Journal of Neuroscience Methods*, vol. 254, pp. 46–59, 2015.

[114] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and M. Joliot, "Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain," *NeuroImage*, vol. 15, no. 1, pp. 273–289, 2002.

[115] S. S. Kessner, E. Schlemm, B. Cheng, U. Bingel, J. Fiehler, C. Gerloff, and G. Thomalla, "Somatosensory Deficits After Ischemic Stroke: Time Course and Association With Infarct Location," *Stroke*, vol. 50, no. 5, pp. 1116–1123, 2019.

[116] J. H. Kim, S. H. Oh, H. J. Jeong, Y. J. Sim, D. G. Kim, and G. C. Kim, "Association between duration of dysphagia recovery and lesion location on magnetic resonance imaging in patients with middle cerebral artery infarction," *Annals of Rehabilitation Medicine*, vol. 43, no. 2, pp. 142–148, 2019.

[117] C. L. Chen, F. T. Tang, H. C. Chen, C. Y. Chung, and M. K. Wong, "Brain lesion size and location: effects on motor recovery and functional outcome in stroke patients." *Archives of physical medicine and rehabilitation*, vol. 81, no. April, pp. 447–452, 2000.

[118] F. Binkofski, R. J. Seitz, T. Hackländer, D. Pawelec, J. Mau, and H. J. Freund, "Recovery of motor functions following hemiparetic stroke: a clinical and magnetic resonance-morphometric study." *Cerebrovascular diseases*, vol. 11, pp. 273–281, 2001.

[119] F. de Shelton and M. J. Reding, "Effect of Lesion Location on Upper Limb Motor Recovery after stroke." *Stroke*, pp. 107–112, 2001.

[120] S. K. Schiemanck, G. Kwakkel, M. W. M. Post, L. J. Kappelle, and A. J. H. Prevo, "Impact of internal capsule lesions on outcome of motor hand function at one year post-stroke," *Journal of Rehabilitation Medicine*, vol. 40, no. 2, pp. 96–101, 2008.

[121] R. Lo, D. Gitelman, R. Levy, J. Hulvershorn, and T. Parrish, "Identification of critical areas for motor function recovery in chronic stroke subjects using voxel-based lesion symptom mapping," *NeuroImage*, vol. 49, no. 1, pp. 9–18, 2010.

[122] B. Cheng, N. D. Forkert, M. Zavaglia, C. C. Hilgetag, A. Golsari, S. Siemonsen, J. Fiehler, S. Pedraza, J. Puig, T. H. Cho, J. Alawneh, J. C. Baron, L. Ostergaard, C. Gerloff, and G. Thomalla, "Influence of stroke infarct location on functional outcome measured by the modified rankin scale," *Stroke*, vol. 45, no. 6, pp. 1695–1702, 2014.

[123] O. Wu, L. Cloonan, S. J. Mocking, M. J. Bouts, W. A. Copen, P. T. Cougo-Pinto, K. Fitzpatrick, A. Kanakis, P. W. Schaefer, J. Rosand, K. L. Furie, and N. S. Rost, "Role of Acute Lesion Topography in Initial Ischemic Stroke Severity and Long-Term Functional Outcomes," *Stroke*, vol. 46, no. 9, pp. 2438–2444, 2015.

[124] R. J. Nudo, "Mechanisms for recovery of motor function following cortical damage," *Current Opinion in Neurobiology*, vol. 16, no. 6, pp. 638–644, 2006.

[125] J. S. Yoo, B. Y. Choi, C. H. Chang, Y. J. Jung, S. H. Kim, and S. H. Jang, "Characteristics of injury of the corticospinal tract and corticoreticular pathway in hemiparetic patients with putaminal hemorrhage," *BMC Neurology*, vol. 14, no. 1, pp. 15–18, 2014.

[126] S. H. Jang, C. H. Chang, J. Lee, C. S. Kim, J. P. Seo, and S. S. Yeo, "Functional role of the corticoreticular pathway in chronic stroke patients," *Stroke*, vol. 44, no. 4, pp. 1099–1104, 2013.

[127] H.-O. Karnath and J. Rennig, "Investigating structure and function in the healthy human brain: validity of acute versus chronic lesion-symptom mapping," *Brain Structure and Function*, vol. 222, no. 5, pp. 2059–2070, 2017.

[128] A. T. DeMarco and P. E. Turkeltaub, "A multivariate lesion symptom mapping toolbox and examination of lesion-volume biases and correction methods in lesion-symptom mapping," *Human Brain Mapping*, vol. 39, no. 11, pp. 4169–4182, 2018.

[129] C.-H. H. Park, N. Kou, and N. S. Ward, "The contribution of lesion location to upper limb deficit after stroke," *Journal of neurology, neurosurgery & psychiatry*, vol. 87, no. 12, pp. 1283–1286, 2016.

[130] E. T. Rolls, C. C. Huang, C. P. Lin, J. Feng, and M. Joliot, "Automated anatomical labelling atlas 3," *NeuroImage*, vol. 206, no. August 2019, p. 116189, 2020.

[131] M. J. Gillies, J. A. Hyam, A. R. Weiss, C. A. Antoniades, R. Bogacz, J. J. Fitzgerald, T. Z. Aziz, M. A. Whittington, and A. L. Green, "The Cognitive Role of the Globus Pallidus interna; Insights from Disease States," *Experimental Brain Research*, vol. 235, no. 5, pp. 1455–1465, 2017.

[132] A. Compston, "Progressive atrophy of the globus pallidus (primary atrophy of the pallidal system). A system disease of the paralysis agitans type, characterized by atrophy of the motor cells of the corpus striatum. A contribution to the function of the corpus striatum," *Brain*, vol. 140, no. 6, pp. 1808–1813, 2017.

[133] M. D. Crutcher and M. R. DeLong, "Single cell studies of the primate putamen - II. Relations to direction of movement and pattern of muscular activity," *Experimental Brain Research*, vol. 53, no. 2, pp. 244–258, 1984.

[134] J. M. Rondina, C. H. Park, and N. S. Ward, "Brain regions important for recovery after severe post-stroke upper limb paresis," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 88, no. 9, pp. 737–743, 2017.

[135] M. Ernst, A. M. Boers, N. D. Forkert, O. A. Berkhemer, Y. B. Roos, D. W. Dippel, A. Van Der Lugt, R. J. Van Oostenbrugge, W. H. Van Zwam, E. Vettorazzi, J. Fiehler, H. A. Marquering, C. B. Majoie, and S. Gellissen, "Impact of ischemic lesion location on the MRS score in patients with ischemic stroke: A voxel-based approach," *American Journal of Neuroradiology*, vol. 39, no. 11, pp. 1989–1994, 2018.

[136] S. H. Jang and H. G. Kwon, "Injury of the cortico-ponto-cerebellar tract in a patient with mild traumatic brain injury," *Medicine*, vol. 96, no. 49, pp. 2–4, 2017.