

https://dspace.org/

# A new method for gesture recognition based on decoding individual finger flexions from surface EMG images

**Ilos, Mishak A**

**2021**
https://hdl.handle.net/10539/33345

Downloaded from a DSpace repository. https://dspace.org/

# A New Method For Gesture Recognition Based On Decoding Individual Finger Flexions From Surface EMG Images

**Mishak A. Illos**




A dissertation submitted to the Faculty of Engineering and the Built Environment, University of the Witwatersrand, Johannesburg, in fulfilment of the requirements for the degree of Master of Science in Engineering.

Johannesburg, October 2020

# Declaration

I declare that this dissertation is my own, unaided work, except where otherwise acknowledged. It is being submitted for the degree of Master of Science in Engineering to the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination to any other university.

Signed this 3 day of November 2020



Mishak A. Illos

# Abstract

People who have suffered a trans-radial amputation are burdened daily by tasks requiring the use of their hands. Myoelectric prosthetic hands provided some level of aid to trans-radial amputees by interpreting electromyographic (EMG) signals into hand movements. However, the actions that the prosthesis can provide are extremely limited by the inability of the device's algorithms to translate complex myoelectric signals into dexterous motions mimicking the human hand. Improvement is required in the area of surface EMG gesture recognition since the current methods of gesture recognition do not provide sufficient classification accuracy over gesture sets that are large enough for adequate prostheses dexterity.

All possible hand gesture can be described by the movements of only five individual fingers. This study will explore a method for hand gesture recognition based on decoding individual finger flexions from high-density surface EMG images. A high-density surface EMG dataset for eight subjects containing eleven gestures was used to test the proposed method against the state of the art gesture recognition method. It is shown that a multi-label convolutional network can recognize the individual finger flexions contained in a set of hand gestures and use the individual finger flexions to infer the eleven gestures to an accuracy of 87.2 %. Additionally, it is shown that eleven gestures can be recognized to an accuracy of 81.1 % when the network was trained on five gestures each containing a single finger flexion.

# Acknowledgements

I would like to thank Mr Abdul-Khaaliq Mohamed and Prof Vered Aharonson for their patience, guidance, and support during the completion of my masters study.

The support of my family has been greatly appreciated. Thank you to my Bob, Lise Essberger for always asking how my Masters is going. Thank you to my Dad, David Illos for cooking. Finally, thank you to my gorgeous Robyn Hingle for helping me set a deadline.

# Contents

**Declaration** **i**

**Abstract** **ii**

**Acknowledgements** **iii**

**Contents** **iv**

**List of Figures** **viii**

**List of Tables** **x**

**1 Introduction** **2**

1.1 The Problem . . . . . 2

1.2 Motivation . . . . . 4

1.3 Dissertation Structure . . . . . 5

**2 Literature Review** **6**

2.1 Introduction . . . . . 6

2.2 sEMG . . . . . 6

2.3 Pattern Recognition in sEMG . . . . . 7

2.3.1 sEMG Electrode Data Capture . . . . . 7

2.3.2 sEMG Data Sampling . . . . . 9

2.3.3 Feature Extraction . . . . . 11

2.3.4 sEMG Pattern Recognition . . . . . 12

2.4 Sequential Classification . . . . . 15

2.5 Simultaneous Classification . . . . . 16

2.5.1 Parallel Classification . . . . . 18

2.5.2 Adapted Multi-Label Classification Algorithms . . . . . 19

2.6 Summary . . . . . 20

**3 Problem Specification** **21**

3.1 Research Questions . . . . . 21

3.2 Research Areas . . . . . 22

3.2.1 Research Area 1 (RA1) . . . . . 22

3.2.2 Research Area 2 (RA2) . . . . . 22

3.2.3 Research Area 3 (RA 3) . . . . . 22

3.2.4 Research Area 4 (RA 4) . . . . . 22

3.3 Constraints . . . . . 22

3.4 Assumptions . . . . . 23

3.5 Hypothesis . . . . . 23

3.6 Expected Outcomes . . . . . 23

**4 Methodology** **25**

4.1 Data . . . . . 25

4.1.1 Participants . . . . . 25
4.1.2 Acquisition Setup . . . . . 25
4.1.3 Acquisition Protocol . . . . . 26
4.1.4 CapgMyo Database . . . . . 26
4.2 Pre-processing . . . . . 27
4.2.1 Reordered Dataset . . . . . 27
4.2.2 Filtering . . . . . 29
4.2.3 Converting to HD-sEMG data to HD-sEMG Images . . . . . 30
4.2.4 Finger Movement Labelling . . . . . 30
4.2.5 Train and Testing Data Separation . . . . . 30
4.3 Procedure . . . . . 32
4.3.1 RA1: Sequential CNN Implementation . . . . . 32
4.3.2 RA1: Validating Sequential CNN on CapgMyo DB-a . . . . . 33
4.3.3 RA2: Test Sequential Classification on CIFF DS . . . . . 33
4.3.4 RA3: CNN Adaptation to Multi-Label . . . . . 34
4.3.5 RA3: Test Multi-label Finger Flexion Classification on CIFF DS 34
4.3.6 RA4: Multi-label Finger Flexion Classification on Unseen CIFF DS gestures . . . . . 35

**5 Results and Discussion** **39**

5.1 RA 1: Validate the classification performance of Sequential CNN on HD-sEMG Images . . . . . 39
5.2 RA2: Sequential Gesture Recognition on CIFF DS . . . . . 41
5.3 RA3: Multi-Label Gesture Recognition on CIFF DS . . . . . 42

5.4 RA4: Multi-Label Gesture Recognition on Unseen CIFF DS Gestures 43

5.4.1 Eleven Gesture Recognition Performance for Best Five Gesture Subset . . . . . 44

5.4.2 Eleven Gesture Performance for Five Single Individual Finger Flexions . . . . . 45

5.5 Answering the Research Questions . . . . . 48

5.5.1 RQ 1 . . . . . 48

5.5.2 RQ 2 . . . . . 49

5.6 Summary . . . . . 49

**6 Conclusion** **51**

6.1 Summary . . . . . 51

6.2 Main Findings . . . . . 52

6.3 Recommendations for Future Work . . . . . 53

**A Sequential Gesture Recognition Results** **61**

**B Multi-Label Gesture Recognition Results** **64**

# List of Figures

2.1 Visual Comparison of Low and High-Density sEMG Electrodes . . . 8

2.2 10 mm Transversal Direction Electrode Shift Collected Using 96 sEMG Channels [1] . . . . . 9

2.3 10 mm Longitudinal Direction Electrode Shift Collected Using 96 sEMG Channels [1] . . . . . 9

2.4 A Comparison Between Sequential Windows and Sliding Window [2] 10

2.5 Classification Accuracy as a Function of the Number of Gestures [3] 16

2.6 Sequential Classification When Dealing With the Multi-Label Problem 17

2.7 Sequential Classification of Grouped and Individual Finger Extensions and Flexions [4] . . . . . 18

2.8 Three Parallel Classifiers, One for Each Colour . . . . . 19

4.1 Finger Flexion Labelling . . . . . 27

4.2 CIFF DS Gesture Set . . . . . 28

4.3 50Hz Powerline Noise in Recorded HD-sEMG Data . . . . . 29

4.4 Convolutional Neural Network Developed in [5] . . . . . 32

4.5 Conversion from Finger Flexion Encoding to Decimal Ground Truth 34

5.1 Sequential Gesture Recognition Accuracy on CapgMyo DBa . . . . . 40

5.2 Confusion Matrix Results for Subjects 1 and 16 from RA1 . . . . . 40

5.3 CapgMyo DBa Gestures 2 and 4 [5] . . . . . 41

5.4 Sequential Gesture Recognition Accuracy as a Function of the Number of Gestures . . . . . 41

5.5 Confusion Matrix Results for Subject 3 on CIFF $\text{Subset}_{5-21}$ . . . . . 42

5.6 Multi-Label Gesture Recognition Accuracy as a Function of the Number of Gestures . . . . . 43

5.7 Multi-Label Gesture Recognition Confusion Matrix . . . . . 45

5.8 Multi-Label Gesture Recognition Confusion Matrix . . . . . 47

5.9 Comparison of SGR and ML-CNN Gesture Recognition Accuracy . . 49

# List of Tables

2.1 EMG Feature Set Performance Using Artificial Neural Networks [6] . 12

2.2 NinaPro Dataset 1 Benchmark Results CNN vs Traditional Classifiers 13

2.3 Single sEMG Colour Frame Classifier Comparison [5] . . . . . 14

2.4 Feature Comparison for Five Gestures & Five Wrist Movements [7] . 15

4.1 Set of Finger Flexions Extracted from CapgMyo DB to Create CIFF DS [8] . . . . . 27

4.2 Inter-Session Correlation of Subject Trails between CapgMyo DBs to form CIFF DS [8] . . . . . 29

4.3 Sequential Encoding Scheme . . . . . 31

4.4 Multi-label Encoding Scheme . . . . . 31

4.5 Five Gesture CIFF $\text{Subset}_5$ . . . . . 36

4.6 Seven Gesture CIFF $\text{Subset}_7$ . . . . . 37

4.7 Nine Gesture CIFF $\text{Subset}_9$ . . . . . 38

5.1 Gesture Recognition Performance Summary for CIFF $\text{Subset}_{5-1}$ . . 44

5.2 Subject 1 Gesture Recognition Performance Report for CIFF $\text{Subset}_{5-1}$ 44

5.3 Subject 8 Gesture Recognition Performance Report for CIFF $\text{Subset}_{5-1}$ 45

5.4 Gesture Recognition Performance Summary for CIFF $\text{Subset}_{5-1}$ . . 46

5.5 Subject 1 Gesture Recognition Performance Report for Single Finger Flexion Gesture Set . . . . . 46

5.6 Subject 2 Gesture Recognition Performance Report for Single Finger Flexion Gesture Set . . . . . 47

A.1 Complete Gesture Recognition Accuracy Results for CIFF $\text{Subset}_9$ . 61

A.2 Complete Gesture Recognition Accuracy Results for CIFF $\text{Subset}_7$ . 62

A.3 Complete Gesture Recognition Accuracy Results for CIFF $\text{Subset}_5$ . 63

B.1 Complete Gesture Recognition Accuracy Results for CIFF $\text{Subset}_9$ . 64

B.2 Complete Gesture Recognition Accuracy Results for CIFF $\text{Subset}_7$ . 65

B.3 Complete Gesture Recognition Accuracy Results for CIFF $\text{Subset}_5$ . 66

# Chapter 1

# Introduction

## 1.1 The Problem

The manufacture and use of limb prostheses dates back hundreds of years. One may conjure up the image of the seafaring pirate, balancing on a wooden pole for a leg, wielding a hook at the lower end of a trans-radial (through the forearm) amputation. When people first survived limb amputations, doctors, engineers and possibly even the victims themselves were inspired to replace the missing limbs with a prosthesis.

The progress in lower limb prostheses has been remarkable with athletes now able to run at speeds close to world records of the unimpaired. Unfortunately, development in the upper limb prosthesis does not compare with that seen in lower limb prostheses. Whilst the materials, mechanical and electronic designs may be capable of enhanced functionality. The methods of device control are lagging behind. In fact, many people with upper limb amputations choose not to wear a prosthesis, because of the limited degrees-of-freedom (e.g. hand open/close) that current commercial prosthetic hand devices offer [9] [10].

Each year, in the UK and Italy, approximately 8700 upper limb amputations are performed. 75% of these amputations are distal to (below) the elbow [11]. An upper limb amputation severely burdens an individual's ability to perform activities of daily living (ADL). There are a vast number of ADL a person may be required to perform using their hand(s), each of which necessitates a specific manipulation of the fingers. The fine motor movements of the hand are incredibly complex and extremely difficult to duplicate when the controlling muscles, nerves, joints and ligaments are lost in an amputation.

To simplify the analysis of ADL requiring the hand, sets of grasps and gestures must be defined for a practical functional hand [12] [13]. A prosthetic hand can be programmed with such a set of grasps or gestures to aid trans-radial amputees in carrying out ADL [14]. The grasps and gestures are prompted by user input, through the detection of myoelectric signals.

The past decade has seen electrically powered prostheses and orthotic devices such as exoskeletons, mainly focus on the use of myoelectric signals for control [15]. Electromyography (EMG) is one such myoelectric signal. EMG is the measurement of electrical signals that are generated when a muscle contracts [16]. EMG data can be collected from different locations on the forearm using a set of electrodes to control a prosthetic hand.

Two types of EMG electrodes exist for measurement, surface EMG (sEMG) and intramuscular (iEMG) [17][18]. In the case of sEMG, electrode pads are placed on the surface of the skin to measure the signals of underlining muscles. Intramuscular EMG electrodes are implanted directly into the muscles under observation. Surface EMG analysis provides a broad platform for engineering and biomedical research into, grasp and gesture recognition without necessitating the same level of ethical consideration required by invasive iEMG electrodes [19]. Additionally, iEMG electrodes have been reported to cause discomfort to the subject [20].

Pattern recognition algorithms can be employed on collected EMG data in order to discriminate between different grasps or gestures. EMG-based models can then be developed to represent each grasp or gesture. The more sophisticated grasp and gesture recognition methods are based on sEMG pattern recognition algorithms [21].

In recent years there has been a rise in the popularity of machine learning (ML) techniques as a method of pattern recognition and classification. The main contributing factors to the increased popularity are; An increase in computer processing power and the ever-increasing availability of the data needed for training networks. Machine learning algorithms have proven to be successful in a wide range of tasks. Machine learning successes in signal processing include Yolo high frame rate object detection [22], Google's acoustic detection of Humpback Whales [23] and Facebook's facial recognition for people with visual impairments [24]. The three ML examples listed utilize a common ML technique called, convolution neural network (CNN) for pattern recognition and discrimination. Additionally, a study by W. Geng et al in the field of sEMG analysis showed that a CNN was highly effective in classifying sEMG data for gesture recognition [5].

Generally, the effectiveness of sEMG recognition models is based on the accuracy to which a set of grasps or gestures can be recognised/classified. For these models to be developed sEMG data needs to be collected for each grasp or gesture type, a grasp or gesture type not contained in the set cannot be recognised [25]. In general, the accuracy of interpreting a user's sEMG data is linked to the amount of sEMG data available (number of sEMG channels, segment length) and the number of grasps or gestures in the set.

Effort into the development of commercial upper limb prostheses has seen production of prosthetic hands which allow for actuation in multiple degrees of freedom (DOF) [26]. Yet the clinical control strategies still only provide control limited to a fixed, predefined set of grasp or gestures [27]. The challenge in providing the prosthesis with accurate human-like dexterity lies in the methods of interpreting the user prompts from the measured sEMG data.

## 1.2 Motivation

A common issue in ML performance is that as the number of classes (types) increases, the classification accuracy (probability of recognition of the type/class) decreases. The same is true when it comes to human hand gesture recognition using ML techniques: There is an excess of possible gestures (classes) required to effectively use ML. Vis-à-vis classification accuracy becomes too low as the size of the set of gestures increase. The difficulty in separating each new gesture increases, since the differences between the features which describe the gestures become subtler.

Numerous studies which have performed EMG data analysis for the recognition of gestures [6][28][29]. These studies have not addressed the potential advantages of developing an sEMG analysis methodology which is able to separate a gesture into the composite individual finger flexions to which it is composed. The focus of these studies has been to achieve higher classification accuracies over a gesture set, rather than determining if signal patterns are shared between similar gestures [30][31][32][25]. Exploiting the similarities between gestures may allow for a smaller gesture sets without limiting the provided control.

When it comes to hand gestures, only five individual fingers (classes) are responsible for all possible hand gestures. Perhaps if we are able to train a neural network to recognise the individual fingers rather than a much larger set of gestures, we could significantly raise the classification accuracy because the class set would be smaller.

The primary assumption is that individual finger states (flexion/non-flexion) can be recognized from sEMG signals.

The study aims to adapt the best currently available pattern recognition method to test if individual finger flexions can be recognised and used to infer the correct gesture [5]. Then to test if this method increases classification accuracy.

A key advantage of an adapted method based on individual finger flexions is that five individual finger flexions could infer up to thirty-two different gestures, provided data existed for all thirty-two gestures. This would reduce the burden faced by trans-radial amputees to adapt to life without a hand.

## 1.3 Dissertation Structure

Chapter 2 presents a review of the key techniques used to develop the sEMG pattern recognition method used to separate gestures into the individual finger flexions. This is followed by an explanation of simultaneous classification and key benefits it has over sequential classification.

Chapter 3 introduces the research questions (RQ's) and the research areas (RA's) required in answering the RQ's. Chapter 3 additionally lists the constraints, assumptions, hypotheses, and expected outcomes of this study.

Chapter 4 describes the CapgMyo database, a high-density sEMG dataset containing eight gestures and twelve individual finger flexions and extensions. The description includes detail on how many subjects participated in the study and how the data was collected. The chapter goes on to describe the testing procedure of both gesture recognition and finger flexion based gesture recognition.

Chapter 5 presents the results of tests performed in the study and provides a discussion of the main findings. The chapter concludes with the answers to study research questions.

Chapter 6 concludes the dissertation with a summary of the main findings and recommendations for future research.

# Chapter 2

# Literature Review

## 2.1 Introduction

A hand amputation renders a considerable burden to the standard of living of the affected person. A hand amputation reduces dexterity and compromises the individual's ability to perform everyday tasks and contribute economically [33]. Simple activities of daily living (ADL) such as tying shoelaces become difficult to accomplish. An actuated prosthesis can provide a platform for people who have suffered a hand amputation to once again be able to carry out ADL [2]. The development of hand prosthesis to aid upper limb amputees over the past decade has seen, extended battery life, weight-saving and improved aesthetic features. However, the clinical control strategies of prosthesis have seen limited development from simple hand open/closed sensing [26]. In the following sections, a review of the methods used to provide control to hand prosthesis will be given. Through this review, it will be shown that the current method of control provided through sEMG gesture recognition is limited by the current classification strategies.

## 2.2 sEMG

Electromyography is a measurement of an electrical signal generated in a muscle when the muscle is in contraction [34]. An action potential (AP) propagates along a nerve from the spinal cord and terminates in the motor end plate, this creates the condition for electrical stimulation of the muscle [34]. The acquisition of the EMG signals is achieved by three types of electrodes; wire, needle and surface [34]. Surface

measurement is the most commonly used since surface EMG (sEMG) is a method of measuring signals non-invasively by placing electrodes on the skin directly above the muscle(s) under interrogation.

Surface EMG analysis provides a platform for biomedical research into human -machine interfacing, prosthesis and orthotic device control, grasp and gesture recognition as well as the clinical diagnosis of neuromuscular disorders. Surface EMG allows for this research without necessitating the same level of consideration for infection and nerve damage required by invasive EMG acquisition techniques [19]. In the section that follows, a broad methodology for recognising muscle movements from sEMG data will be introduced.

Control of prosthetic hands is provided through the analysis of myoelectric signals or via body-powered prompts [35]. Myoelectric control commonly uses EMG signals since the information contained in the signals is localised to the area of muscle under inspection. EMG provides muscle activation patterns which allow for movements to be interpreted [36].

## 2.3 Pattern Recognition in sEMG

Pattern recognition algorithms provide a means of separating input data into respective classes based on relevant features. The past two decades have seen a trend toward sEMG pattern recognition of hand movements and gestures for sEMG analysis and control[26] [37]. The recognition of the user's muscle activation patterns attributed to different movements provides the platform for multifunction control of prostheses.

Pattern recognition methodology for gesture classification includes; capturing raw sEMG data, windowing the data for each channel, extracting features and finally providing a classification [28].

In the subsections to follow a critic of each of the prior mentioned gesture classification process will be discussed.

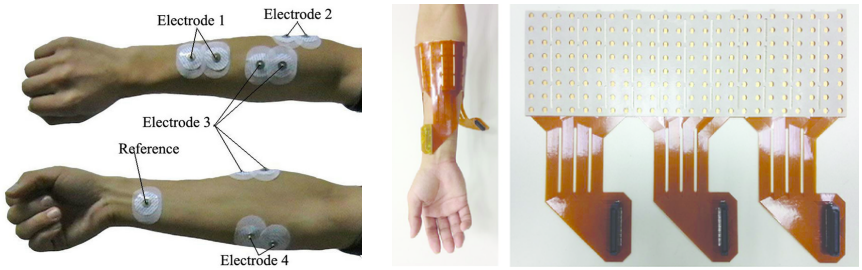
### 2.3.1 sEMG Electrode Data Capture

In the field of sEMG pattern recognition, lower count sEMG electrode channels have been favoured over high density-sEMG (HD-sEMG) [38] [39]. The increase in

the number of sEMG channels increases the complexity of the data capture [39]. Each sEMG channel additionally requires several pre-processing steps before the sEMG data can be fed into a movement classifier [28] [40]. To improve the quality of classification, using minimal electrode channels, research into electrode placement has sought the optimal location for sEMG electrodes [41] [42].

Pattern recognition algorithms reliance on specialised electrode placement does, however, create a susceptibility to misclassification in the event of electrode displacement [37]. Electrode displacement is the result of either electrode shift or slight deviations in electrode placement between sessions. During pronation or supination of the forearm, the underlining muscle location relative to the placement of electrodes on the skin changes. Electrode shift is hence unavoidable when using sEMG.

In addition to electrode displacement, low count sEMG provides less information on muscle activation patterns than HD-sEMG. Low count sEMG provide less information since there is less muscle surface area coverage [38]. Figure 2.1a demonstrates the sparsity of the low count electrode coverage as compared to 192 channel HD-sEMG seen in fig 2.1b.




(a) Low count sEMG electrodes [32] (b) HD-sEMG electrodes [1]

Figure 2.1: Visual Comparison of Low and High-Density sEMG Electrodes

HD-sEMG, when used in conjunction with a specialised classifier and training data (includes pronation and supination of the forearm) yields higher classification accuracy than low count sEMG [39]. In fig 2.2 the effect 10 mm transversal direction electrode shift has on measured muscle activation patterns of ninety-six sEMG channels is seen [1]. The overall sEMG pattern for this gesture remains similar, the bottom left and right corners have the highest sEMG amplitudes. This is similar for the 10 mm longitudinal direction electrode shift. The bottom left and right corners have the highest sEMG amplitudes (fig 2.3).

To summarise, HD-sEMG provide a better platform for gesture classification than low

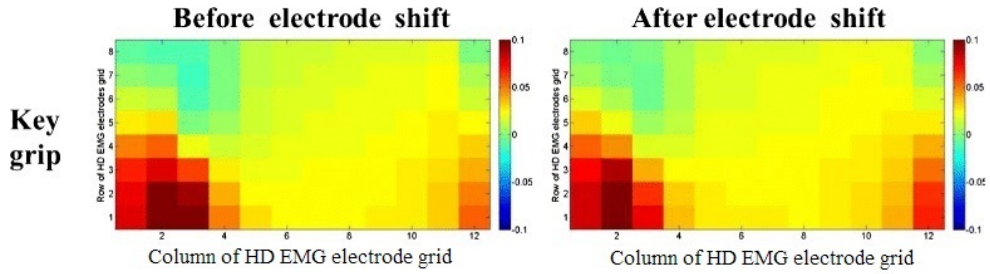



Figure 2.2: 10 mm Transversal Direction Electrode Shift Collected Using 96 sEMG Channels [1]

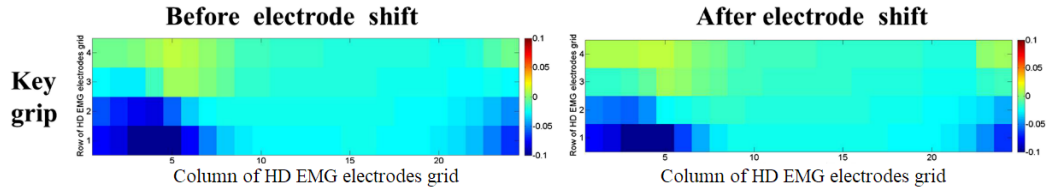


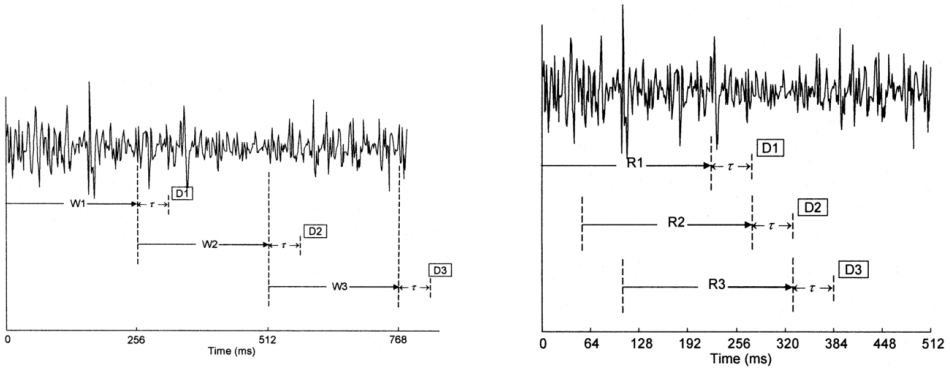

Figure 2.3: 10 mm Longitudinal Direction Electrode Shift Collected Using 96 sEMG Channels [1]

count sEMG for the following reasons; 1. HD-sEMG provides more muscle activation information because of the greater muscle surface coverage. 2. HD-sEMG data is more robust against electrode displacement caused by pronation and supination of the forearm. The section that follows will introduce and explore the benefits of two different methods of sEMG data sampling, namely windowing and imaging.

### 2.3.2 sEMG Data Sampling

There are two main methods for sampling sEMG data, windowing and imaging. Windowing is the process of segmenting the raw sEMG data into a finite sequence of data points collected over a predefined time interval [2]. The window size is subject to the sample rate and the number of required data points needed for gesture recognition. There are two types of windowing, sequential windowing in which one complete window follows the previous window (see fig 2.4a. The second type is a sliding window in which each window overlaps by the window length less the step size $\tau$ (see fig 2.4b). The time to process the data in the window must not exceed $\tau$, since this will result in a cumulative delay. Sliding windows additionally act to smooth the transition between windows, thereby reducing the effects of artifacts and noise [43].

The process of windowing sEMG data introduces a delay between initiation to




(a) Sequential windows (b) Sliding windows

Figure 2.4: A Comparison Between Sequential Windows and Sliding Window [2]

perform the movement and detection of the movement. The introduced delay is calculated as the sum of the window size plus processing time [40]. Latency in prediction which is greater than 100 ms negatively affects the prostheses performance and ultimately lead to amputees discontinuing the use of the prostheses [9] [44]. Smaller window sizes are possible ($<$ 100 ms), however reducing window size does hamper classification accuracy [45].

A newly developed technique of sEMG imaging has been used for gesture recognition. The technique uses the most recently captured sEMG channel values and maps each channel value to a corresponding pixel in an image [5]. The size of the sEMG image is a function of the number of electrodes channels used.

Mapping each sEMG channel to a specific pixel encodes the temporal-spatial patterns to the image. This is particularly useful for sEMG analysis because it links specific muscle contractions to regions of the sEMG image.

Imaging of sEMG data still requires windowing the sEMG channels before each sEMG channel can be mapped to a pixel, however, EMG imaging has been shown to achieving high gesture recognition accuracy with the window lengths $<<$ 100 ms [46] [47].

The next section will explore methods of extracting pattern features from sEMG data for the use in gesture recognition.

### 2.3.3 Feature Extraction

Historically pattern recognition algorithms have relied on a vector of features which describe the raw sEMG data. The ability of a pattern recognition algorithm to correctly identify patterns depended almost entirely on which features were extracted [47]. The sEMG features are extracted from three broad categories; time domain, frequency domain and time-frequency domain.

Time-domain features have been extensively used in sEMG pattern recognition because of the relative simplicity and more generalised performance [2]. Features from the time-domain are inherently base on signal amplitude. The amplitude component of time-domain features is particularly influenced by factors such as electrode location, tissue thickness, muscle contraction velocities and distribution of motor units in the muscle fiber [28]. The more extensively used features are mean absolute value (MAV) and root mean square (RMS). The values of MAV and RMS are calculated as follows:

$$MAV = \frac{1}{N}\sum_{i=1}^{N}|x_1| \tag{2.1}$$

$$RMS = \sqrt{\frac{1}{N}\sum_{i=1}^{N}{x_1}^2} \tag{2.2}$$

where $x_i$ are the signal samples, and N the number of samples in the data segment considered.

The frequency domain is more commonly used for the study of muscle fatigue through the inspection of changes in the firing rate of recruited muscle units [48]. Features such as power spectral density (PSD) and mean frequency are extracted from the frequency domain of raw sEMG data, however, the window size does affect frequency resolution and frequency leakage [28]. Moreover, once a signal has undergone frequency analysis the time domain information is lost.

Time-frequency analysis features can be extracted using short-time Fourier transform (STFT) or wavelet transform (WT), which contain information for both time and frequency [49]. The difference between STFT and WT is in the way the time-scale axis is portioned [28]. The variable portioning ratio of the WT has allowed for a family of WT coefficient which has extended the use-case to de-noising and isolating coordinated muscle activities [50].

Table 2.1: EMG Feature Set
Performance Using
Artificial Neural Networks [6]

| **Feature Set** | **Average%** |
|---|---|
| Wavelet DAU20 | 59.6% |
| Frequency domain | 62.5% |
| Wavelet CH | 63.3% |
| Wavelet BL | 65.8% |
| Wavelet DAU4 | 66.2% |
| Time-domain | 78.3% |

In a review of EMG signal analysis techniques, M. Raez et al. presented a performance comparison of the following features; Time domain, Frequency domain and Wavelet coefficients using an artificial neural network [6]. The results are summarised in Table 2.1, we see that the time domain features provide the best classification accuracy, which is supported by the literature [37][51].

To summarise, the literature suggests that time-domain features provide the best opportunity for sEMG pattern recognition. The recognition of sEMG patterns is discussed in the section to follow.

### 2.3.4 sEMG Pattern Recognition

Data classification is the broad subject which encapsulates a large variety of algorithms. In the field of sEMG pattern recognition, a classifier is a tool used to separate the different sEMG patterns. These algorithms are designed to make a classification decision based on the patterns derived from the data features presented. The performance obtained from any given classifier is highly dependent on the structured input data. Only data information (e.g feature space) that is relevant to the classification/recognition process should be presented. This means that for effective classification the input data information needs to be well defined.

The most commonly used classifiers in sEMG gesture classification are linear discriminate analysis (LDA), support vector machines (SVM) and K-Nearest Neighbours (K-NN) [52]. Raw sEMG data is rarely used as an input but rather a selection of features are extracted by performing a mathematical operation on the raw data. The relevance of extracted features differs with respect to sEMG patterns produced

by the individuals and the severity of the amputation [53] [54]. The differences in sEMG patterns affect which features are most effective in classification. The typical approach to ensuring that the most relevant features are used from individual to individual is to implement a feature reduction technique such as principal component analysis (PCA) [55].

In recent years convolutional neural networks (CNN) have been shown to be a highly effective machine learning technique for image classification. A CNN is best suited to image classification since the network is spatially invariant, meaning that classification accuracy is less affected by the location of a pattern in a 2D space [56]. The spatial invariance of a CNN provides a key advantage over other neural networks in sEMG pattern recognition since a CNN is less affected by electrode shift [57]. As a result of the spatial invariance and feature extraction through convolutional layers, a CNN was shown to be effective in hand gestures classification using sEMG data [5] [47] [57]. In the study by W.Geng et al. it was shown that eight isometric and isotonic gestures could be classified with an accuracy of 86.4% using only a single colour image of sEMG data (equivalent to 3 ms) [5]. Where a colour image has a resolution of 16x8 px, 128 electrodes are organised into the shape (16x8) and layered three times $t_n, t_{n+1}$ and $t_{n+2}$ to achieve an RGB (red, green, blue) colour space.

Table 2.2: NinaPro Dataset 1 Benchmark Results
CNN vs Traditional Classifiers

| **Classifier** | **Features** | **window length (ms)** | **Average Accuracy** |
|---|---|---|---|
| LDA [58] | mDWT | 200 | 59 |
| K-NN [58] | RMS | 200 | 65 |
| SVM [58] | RMS, TD, HIST, mDWT | 200 | 70 |
| Random Forest [58] | RMS, TD, HIST, mDWT | 200 | 75.32 |
| CNN [5] | - | 10 | 65.1 |

In Table 2.2 the classification performance results are shown for CNN, Random Forest, SVM, K-NN and LDA on the NinaPro Dataset 1. The NinaPro dataset is a scientific benchmark dataset for gesture classification [58]. Dataset 1 of NinaPro contains fifty-two different gestures performed by twenty-seven able subjects. The features used for training and testing the Random Forest and SVM classifiers were root mean squared (RMS), time-domain statistics (TD), histogram (HIST) and marginal discrete wavelet transform (mDWT). The features that resulted in the best accuracy for K-NN and LDA were RMS and mDWT respectively. The CNN

was trained and tested using a single colour image of sEMG data (equivalent to 10 ms) of resolution 1x10 px. The best performing classifier was the random forest, however, the CNN was able to achieve 65.1% despite relative window length of 10 ms compared to 200 ms (a twentieth of the data) for the random forest.

Motivated by the results on the NinaPro dataset, W. Geng et al. compared six classifiers to test if it was possible to classify gestures without performing feature extraction [5]. The data used for the set was a 128-channel, HD-sEMG data organised into a 16 by 8, three channel colour frame. The classifiers tested were; Multilayer perceptron (MLP), Random Forests, KNN, SVM, LDA and CNN. It was observed that patterns could be recognised directly from sEMG gesture data with MLP, Random Forests and CNN without having to extract features. The results for this test are shown in Table 2.3.

Table 2.3: Single sEMG Colour Frame Classifier Comparison [5]

| **Classifier** | **Average Accuracy %** |
|---|---|
| SVM | 14 |
| LDA | 16 |
| KNN | 40 |
| MLP | 60 |
| Random Forest | 61 |
| CNN | 86.4 |

The best performing classifier on HD-sEMG image was the CNN. The CNN on average was 25.4% better classification accuracy than the next best classifier, the random forest.

In the study by H. Chen et al. three-time domain features (MAV, SSC (Slope Sign Changes), and ZC) were compared to the features extracted by a CNN [7]. The features were extracted for five wrist and five hand gestures and then classified using an SVM classifier. H. Chen et al. used a sliding 300 ms sliding window on 16 channel sEMG to create a grayscale image with a resolution of 300x16 px as input for the CNN. The averaged classification accuracies over all subjects and gestures examined in the study by Chen et al. are compared in Table 2.4 [7].

Table 2.4: Feature Comparison for Five Gestures & Five Wrist Movements [7]

| **Hand Action** | **Feature** | **Average Accuracy %** |
|---|---|---|
| Five Wrist Movements | ZC | 55.73 |
| | SSC | 58.21 |
| | MAV | 62.03 |
| | CNN | 73.79 |
| Five Gestures | ZC | 33.41 |
| | SSC | 34.04 |
| | MAV | 41.12 |
| | CNN | 49.88 |

In Table 2.4 it is shown that a CNN feature extraction outperforms MAV by 11.76% over five wrist movements and by 8.76 % for the five hand gestures.

The literature shows that a CNN often provides higher classification accuracies over commonly used classifiers such as K-NN, random forest, MLP and SVM. The improved performance of a CNN on gesture classification is related to the features extracted through convolution and spatial invariance of CNN [5]. The CNN performance was additionally improved when used in combination with HD-sEMG imaging.

## 2.4 Sequential Classification

The previously mentioned methods of pattern recognition are fundamentally sequential in nature, only able to predict a single gesture per time window or image. Furthermore, the sEMG patterns which can be recognised are limited to a pre-defined set of hand gestures for which the model was constructed to recognise. This method of sequential pattern recognition hence provides dexterity which is limited to this finite set of movements.

In the case of control for a prosthetic hand with two degrees of freedom (DOF), wrist flexion/extension (WF/WE), hand open/closed (HO/HC), only a single degree of freedom can be identified at a time. To mitigate this problem researchers have expanded the set to include additional combined movements. This, in turn, increases the number of patterns that need to be recognised. Moreover, by expanding the

gesture recognition algorithm to consider the angle to which a finger flexion is performed further complicates the problem.

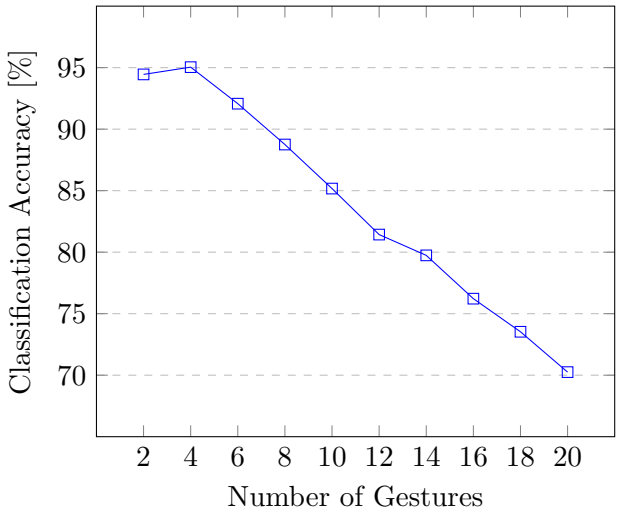



Figure 2.5: Classification Accuracy as a Function of the Number of Gestures [3]

In fig 2.5 the average accuracy of prediction for an increasing number of gestures is plotted. The classifier used linear discriminate analysis (LDA) for feature reduction and support vector machine (SVM) for prediction.

Research analysis revealed a trend; as the number of gestures studied increased, the sEMG recognition accuracy began to decrease, as depicted in fig 2.5. The apparent reason for this inverse relationship is that as the number of investigated gestures increase, the differences in the sEMG patterns of the gestures become more subtle and therefore more difficult to separate. As previously mentioned in chapter one, section two, this is a common problem with sequential pattern classification. The investigator attempts to define more and more classes (i.e. increasing the output dimensionality) without increasing the "pattern generating data" (useful features) from the available input information.

Improving the controllability of hand prostheses requires overcoming the limitations of the sequential classification approach. Simultaneous classification-based control which is discussed in the section to follow provides a possible solution.

## 2.5 Simultaneous Classification

Simultaneous classification is a method of providing multiple classifications for a given set of data points. In the field of machine learning, this is referred to as the

multi-label classification problem. Multi-label classification occurs when multiple outputs can be classified (labelled) for a single input data vector. Similarly, in the case of myoelectric control, it is the equivalent of breaking down a complex movement into each individual DOF.

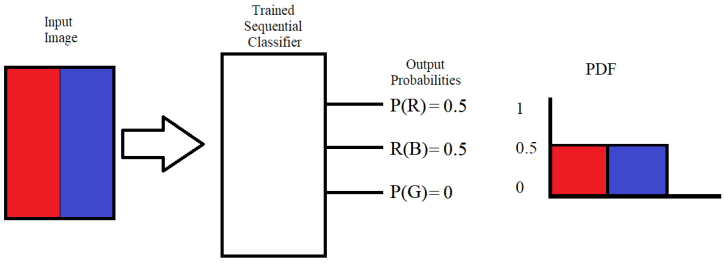



Figure 2.6: Sequential Classification When Dealing With the Multi-Label Problem

A simplified example of the multi-label problem is shown in fig 2.6, a classifier must label the input image as “red”, “blue” or “green”. In the case of multiple true labels (the image coloured red and blue), a sequential classifier will distribute the probability such that it is shared evenly over all the true outcomes. The probability distribution function (PDF) in fig 2.6 shows this. A softmax function is typically used for sequential classification.

Softmax: Maps $n$ outputs ($x$) from real numbers to probability distribution.

$$softmax(x_i) = \frac{e^{x_i}}{\sum_j^n e^{x_i}} \tag{2.3}$$

The solution in sequential classification to deal with multiple true outcomes has been to add a new class for “blue & red”, “ red & green”, etc. however as discuses in section 2.4 (Sequential Classification), increasing the output dimensionality reduces average classification accuracy.

In fig 2.7 the classification accuracy of twelve different finger movements classes is compared. The movements are labelled as flexion (“f”) and extension (“e”), with the subscript indicating which finger, 1 for the thumb and 5 for the little finger. The case of combined finger extension (e345) and flexion (f345) can be defined as a multi-label problem, by allowing for the case where e1 & e2 & e3 are all true. In the study by F. Tenore et al. the combined finger movements (e345 & f345) were instead considered as new separate classes, resulting in an increase in the number of classes [4]. Incidentally considering the combined finger movements (e345 & f345) as classes resulted in cross-class confusion (indicated by red arrows). The cross-class

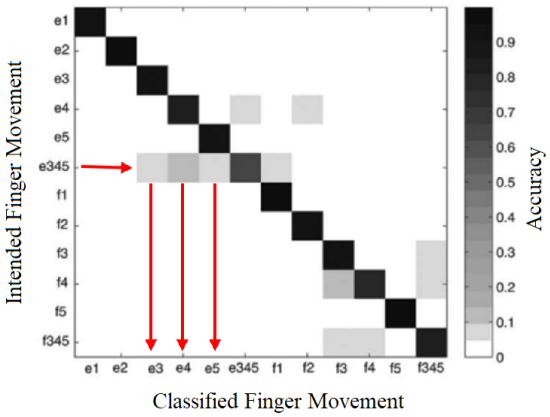



Figure 2.7: Sequential Classification of Grouped and Individual Finger Extensions and Flexions [4]

confusion between e345 and the individual extensions of e3, e4 and e5 suggest that there may be a relationship between sEMG signals for individual finger movements and the combined finger movement.

### 2.5.1 Parallel Classification

The multi-label problem introduced in fig 2.6, which required a classifier to label a red and blue input image as “red”, “blue” or “green” can be solved using parallel binary classifiers (as seen in fig 2.8). A parallel binary classifier uses the combination of two or more separately trained binary classifiers, each binary classifier is responsible for detecting the presence of a class. The coloured image is classified by using a binary classifier for each colour (ie. P(blue) vs P(!blue)). This approach allows for reduced output dimensionality and label independence.

In the study by A. Young et al., the simultaneous classification control approach was used in identifying hand and wrist movements from sEMG [35]. The movements for the hand and wrist were treated as discrete components of a combined hand-wrist movement. It was shown that by using three parallel binary classifiers it was possible to deconstruct a combined hand-wrist movement into a wrist flexion/extension (WF/WE), hand open/closed (HO/HC) and no movement. This approach provides control information for five movements but reduced the output dimensionality of the sEMG classifier from five to three.

Similarly, in work done by J. Baker et al., parallel classifiers were used on implantable myoelectric sensors (IMES) data [51]. This was shown to be effective in decoded

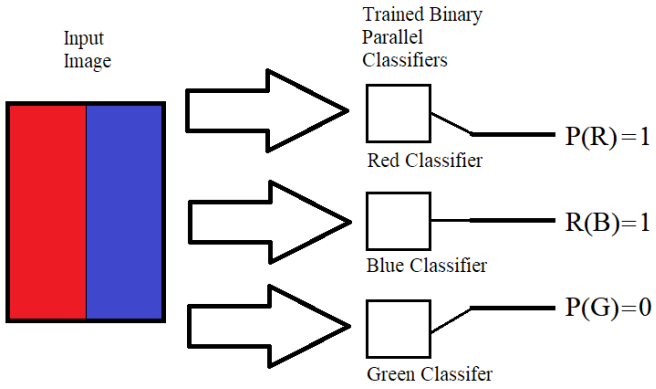



Figure 2.8: Three Parallel Classifiers, One for Each Colour

individual flexions of the thumb(T), index(I), and middle(M) finger from combined finger flexions of a rhesus monkey. The combined finger flexions studied in [51] were TI, TM, IM and TIM. A hand gesture could similarly be decoded into single finger flexions and extensions, provided that sEMG patterns for single finger actions are decodable from that of a grouped finger action.

### 2.5.2 Adapted Multi-Label Classification Algorithms

The common strategy in multi-label classification is to use multiple parallel binary classifiers (as seen in subsection 2.5.1). When using parallel classifiers, the dependency between classes is not modelled, since then training each binary classifier, the classes are treated as independent from one another [59]. Hence correlations between classes that exist in the training data is ignored. This results in poor classification since conflicting classes can be grouped [60]. A hypothetical example of this would be a parallel classifier to recognize bird species and species colour. An example of a conflicting result would be classifying a black swan as a purple swan, since a swan cannot be purple. To best model the cross-class dependence of classes, a single classifier should be used, and the output layer adapted to support multi-label classification. In the field of machine learning, this classification technique is referred to as “adapted multi-label classification”.

The classification of text is an example of a typical multi-label problem, for any given text could be assigned multiple labels (ie. news article could be classed as politics and finance). In [61] the XML -CNN (eXtreme Multi-Label) was designed to classify multi-label texts from a data set which consisted of over 2 million documents and 500,000 labels. XML-CNN was adapted to support multi-label classification by

utilising binary cross-entropy loss over sigmoid output. A sigmoid (Eq. 2.4) output allows for multi-label classification since as opposed to sequential classification the probability for each label is not distributed over the set.

Sigmoid: Real value non-linearly mapped to the interval between [0,1]

$$\sigma(x) = \frac{1}{1 + e^x} \tag{2.4}$$

A CNN used for gesture classification could be adapted to support multi-label classification using a similar method used in XML- CNN.

## 2.6 Summary

An examination of the literature in the field of gesture recognition revealed that a CNN in combination with HD-sEMG imaging yields the best classification accuracy for gesture recognition. However, despite the relatively high classification accuracies obtained when using a CNN, as the number of gestures increase, the classification accuracy reduces.

To improve gesture classification a method is required that reduces the rate at which classification accuracy drops as the number of gestures increase.

Simultaneous classification provides a platform to limit the number of gesture classes to only the degrees of freedom under inspection, rather than having to add new classes for each new gesture.

The classifier that provides the best opportunity for multi-label adaption, as well as extracting the sEMG features that relate similar movements in hand gestures, is a convolutional neural network using high density-sEMG images.

# Chapter 3

# Problem Specification

In this chapter, the research questions (RQs) and research areas (RAs) for this study are presented in 3.1 and 3.2. The constraints and assumptions are detailed in sections 3.3 and 3.4, respectively. The hypotheses for the study are defined in sections 3.5, and section 3.6 lists the expected outcomes of the study.

## 3.1 Research Questions

The proposed study aims to answer the following research questions;

RQ 1. Using a 128 channel sEMG dataset, can gesture recognition accuracy be improved from the current state of the CNN gesture recognition method by using a multi-label CNN to recognize five individual finger flexions and infer the correct gesture?

RQ 2. Using a 128 channel sEMG dataset, to what accuracy can eleven gestures be recognized by the multi-label CNN trained on only five gestures?

In the RQs, gesture recognition accuracy refers to the comparison between ground truth gesture, the correct recognition of that specific gesture from HD-sEMG data. The gesture recognition accuracy indicates the performance of the gesture recognition methodology.

## 3.2 Research Areas

To answer the research questions, the following research areas have been defined.

### 3.2.1 Research Area 1 (RA1)

Validate the performance of the state of the art sequential CNN for HD-sEMG gesture classification described in the study by W. Geng et al. on eight isometric and isotonic hand gestures [5].

### 3.2.2 Research Area 2 (RA2)

Determine the rate at which classification accuracy decreases as the number of gestures classes increase, for the sequential CNN gesture classification. This was done using five, seven, nine and eleven gestures. This created a baseline to which the multi-label CNN is compared.

### 3.2.3 Research Area 3 (RA 3)

Develop a multi-label CNN capable of simultaneously recognising multiple finger flexions, and test the accuracy of this network on five, seven, nine and eleven gestures.

### 3.2.4 Research Area 4 (RA 4)

Using the multi-label CNN capable of simultaneously recognising multiple finger flexion developed for RA 3, test the gesture recognition accuracy of the multi-label CNN across eleven gestures when trained on a selection of five gestures.

## 3.3 Constraints

The following constraints have been identified;

1. The gestures under inspection are limited to those recorded in the database provided by W. Geng et al in [5].

2. The database of gestures does not provide any information indicating the force exerted for either individual finger flexions or gestures [5].

3. The database of gestures does not provide any information about the individual finger joint angles during flexion [5].

## 3.4 Assumptions

The following assumptions have been identified;

1. For each subject, sEMG pattern commonalities exist between each finger flexion and the gestures containing that same finger flexion.

2. The sEMG patterns related to each finger flexion and each gesture are reproducible for each subject between trials.

## 3.5 Hypothesis

The following hypotheses have been defined.

1. The gesture recognition accuracy of eleven gestures can be improved by decoding the individual finger flexions contained in each gesture using a multi-label convolutional neural network trained on gestures with individual finger flexion labels.

2. A multi-label convolutional neural network trained on five gestures comprised of individual finger flexions will allow for the recognition of eleven gestures.

## 3.6 Expected Outcomes

The expected outcomes included;

1. Gesture recognition by means of individual finger decoding using multi-label CNN on HD-sEMG data.

2. Gesture recognition of gestures unseen during training by means of individual finger decoding using multi-label CNN on HD-sEMG data. This has not been attempted in previous studies, to the author's best knowledge.

# Chapter 4

# Methodology

## 4.1 Data

The data used in this study is found in the CapgMyo database [8], which was first published in 2016 [5] and has since been used in sEMG studies by R. Khushaba et al. and W. Wei et al. [62][25]. The database consists of sEMG data collected using 8x16 electrode array to record hand gestures performed by twenty-three participants. The sections to follow provide further detail on the properties of the database. Ethical clearance for the use of the CapgMyo database has been obtained from Wits, Human Research Ethics Committee (Medical).

### 4.1.1 Participants

The twenty-three participants were healthy, able-bodied subjects ranging in age from twenty-three to twenty-six years. Each subject was paid to perform a set of gestures while wearing a non-invasive acquisition device [8].

### 4.1.2 Acquisition Setup

A non-invasive wearable device developed by Y. Du was used to collect HD-sEMG data [8]. The wearable device contained eight acquisition modules. Each module consisted of an 8x2 array of differential electrodes. The diameter of each electrode was 3 mm with an inter-electrode distance of 7.5 mm horizontally and 10.05 mm

vertically. The electrodes were silver wet and covered with conductive gel, yielding a contact impedance of less than 3 kΩ.

The wearable device developed by Y. Du et al. was fixed onto the right forearm by adhesive bands [8]. The first of eight modules were placed on the extensor digitorum communis muscle at the height of the radio-humeral joint. Each of the subsequent modules were placed clockwise around the subject's forearm, creating an 8x16 electrode array. The sEMG signals were acquired from the wearable device using a 16-bit A/C conversion sampling at 1000 Hz, and the sEMG signals were then band-pass filtered at 20-380 Hz.

### 4.1.3 Acquisition Protocol

The CapgMyo database provided by [8] was collected by first making subjects watch a tutorial video to acquaint themselves with the experiment. Each subject skin was cleaned with rubbing alcohol prior to electrode placement. During the acquisition, each subject was seated with their forearm rested on a desk. The subjects were instructed to copy the gestures displayed on a screen with their right hand.

### 4.1.4 CapgMyo Database

The CapgMyo database contains three movement categories (1) eight isometric and isotonic hand gestures; (2) Twelve finger movements; (3) Two hand gestures used for the estimation of maximal voluntary contraction (MVC) force.

CapgMyo DB-a and DB-b contain eight isometric and isotonic hand gestures. CapgMyo DB-a was collected from eighteen participates in one recording session. CapgMyo DB-b was collected from twenty participates over two sessions.

The CapgMyo DB-c is a database containing the individual flexion and extension of each finger as well as two max force gestures [8]. The HD-sEMG data was collected in one session using ten participates. In each of the three CapgMyo databases (a, b and c) each participant performed each movement ten times. Each movement was held for 3 to 10 seconds.

## 4.2 Pre-processing

### 4.2.1 Reordered Dataset

In order to achieve the research aims presented in chapter three and ultimately answer the research questions, a dataset of gestures containing distinct individual finger flexions was assembled for this study from a selection of gestures in CapgMyo DB. This dataset (DS) was called CIFF (Containing Individual Finger Flexion). The fingers in fig 4.1 are colour coded to relate to Table 4.1.

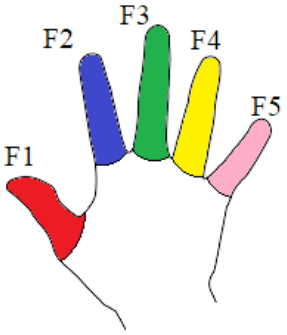



Figure 4.1: Finger Flexion Labelling

Table 4.1: Set of Finger Flexions Extracted from CapgMyo DB to Create CIFF DS [8]

| | | | Individual Finger Flexions: Multi-Label Classes | | | | |
|---|---|---|---|---|---|---|---|
| | CapgMyo DB | CIFF DS Index | F1 | F2 | F3 | F4 | F5 |
| Gestures: Sequential Classes | a&b 1 | 1 | | x | x | x | x |
| | a&b 2 | 2 | x | | | x | x |
| | a&b 3 | 3 | | | | x | x |
| | a&b 6 | 4 | x | x | x | x | x |
| | a&b 7 | 5 | x | | x | x | x |
| | a&b 8 | 6 | | | | | |
| | c 1 | 7 | | x | | | |
| | c 3 | 8 | | | x | | |
| | c 5 | 9 | | | | x | |
| | c 7 | 10 | | | | | x |
| | c 11 | 11 | x | | | | |

The second column of Table 4.1 indicates the name of the gesture in the CapgMyo DB eg. (11 c is from CapgMyo BD-c gesture 11). Column three is the assigned CIFF DS index. Columns four through eight represent the fingers in flexion, starting at the thumb (F1) and ending at the little finger (F5) (as seen in fig 4.1).

Fig 4.2 provides an illustration of all the gestures contained in CIFF DS. Beneath each gesture seen in fig 4.2 is the CIFF DS index and the corresponding CapgMyo database and gesture number. The CIFF DS gesture set is hence the amalgamation




Figure 4.2: CIFF DS Gesture Set

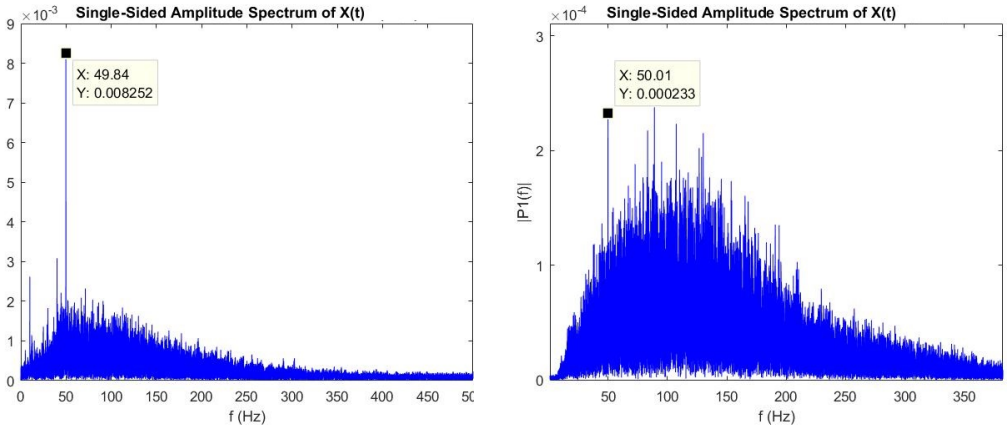
of three datasets collected during different sessions. In table 4.2 the subject IDs are for each database are linked in each row. When a subject was included in both both CapgMyo DB a and b the data for CapgMyo DB a was included in CIFF DS.

Table 4.2: Inter-Session Correlation of Subject Trails between CapgMyo DBs to form CIFF DS [8]

| Db-a | Subject ID | Db-b | Subject ID | Db-c | Subject ID |
|---|---|---|---|---|---|
| x | 1 | x | 1 | x | 2 |
| - | - | x | 5 | x | 3 |
| x | 4 | x | 7 | x | 4 |
| - | - | x | 15 | x | 5 |
| x | 9 | - | - | x | 6 |
| x | 11 | - | - | x | 8 |
| x | 16 | - | - | x | 9 |
| x | 18 | x | 19 | x | 10 |

### 4.2.2 Filtering

The frequency analysis of the raw sEMG signals for the gestures in the CapgMyo DB was shown to contain 50 Hz powerline noise. An example of the 50 Hz noise is illustrated in both the thumb up (CIFF DS - 1) gesture and little finger flexion (CIFF DS - 10), see fig 4.3.




(a) Thumbs Up Gesture (b) Little Finger Flexion

Figure 4.3: 50Hz Powerline Noise in Recorded HD-sEMG Data

The HD-sEMG data contained in the CIFF DS and the eight isometric and isotonic hand movements found in CapgMyo DB-a were band-stop filtered between 45-55 Hz to remove powerline interference as recommended in the study by W. Geng et al. [5].

### 4.2.3 Converting to HD-sEMG data to HD-sEMG Images

The 128 sEMG channels used to record the muscle activation signals are linearly mapped from the interval [-2.5mV, 2.5m] to [-1,1]. The 128 sEMG mapped electrode channels were then arranged into a matrix ($M_{EMG}$) of size 16 by 8 and depth $N$ ($M_{EMG}[16:8:N]$), where N is the number of samples for a trial. The 16 by 8 matrix is converted to a multi-channel image using the equation 4.1:

$$\text{Image}(n) = \begin{cases} \text{Channel 1} & M_{EMG}[16:8:n] \\ \text{Channel 2} & M_{EMG}[16:8:n+1] \\ \text{Channel 3} & M_{EMG}[16:8:n+2] \\ \text{Channel 4} & \frac{|Channel1|+|Channel2|+|Channel3|}{3} \end{cases} \tag{4.1}$$

Once converted to images the CapgMyo DB-a and CIFF DS produced 9980 sEMG images, per gesture, per subject.

### 4.2.4 Finger Movement Labelling

The encoding scheme for the sequential and multi-label classification methods is shown in Table 4.3 and 4.4 respectively. Table 4.3 is the sequential gesture encoding for the eight gestures found in CapgMyo DB-a and the eleven gestures found in CIFF DS. The number of bits is equal to the number of movement classes in each set to be classified, only a single bit can be set to one for any movement class.

Table 4.4 is the multi-label finger flexion encoding for the eleven gestures in CIFF DS. The number of bits is equal to the number of individual finger flexions (five, one bit for each finger). Each bit corresponds to a finger flexion and the bit is set to one if the gesture involves the flexion of that finger.

### 4.2.5 Train and Testing Data Separation

Training and test data was required for the two network architectures, Sequential Gesture Recognition CNN and Multi-Label CNN. The details of the two CNN architectures are described in the procedure section that follows. The sequential-CNN was trained and tested on two data sets, the eight gestures found in the CapgMyo DB-a and the set of eleven gestures found in CIFF DS (see table 4.3 ).
The multi-label CNN was trained and tested on the detection of finger flexions using

Table 4.3: Sequential Encoding Scheme

| CapgMyo DB-a Gesture Classification | Encoding | CIFF DS Gesture Classification | Encoding |
|---|---|---|---|
| 1 | 00000001 | 1 | 00000000001 |
| 2 | 00000010 | 2 | 00000000010 |
| 3 | 00000100 | 3 | 00000000100 |
| 4 | 00001000 | 4 | 00000001000 |
| 5 | 00010000 | 5 | 00000010000 |
| 6 | 00100000 | 6 | 00000100000 |
| 7 | 01000000 | 7 | 00001000000 |
| 8 | 10000000 | 8 | 00010000000 |
| | | 9 | 00100000000 |
| | | 10 | 01000000000 |
| | | 11 | 10000000000 |

Table 4.4: Multi-label Encoding Scheme

| CIFF DS Index | Finger Flexion Encoding | Decimal Value |
|---|---|---|
| 1 | 11110 | 30 |
| 2 | 11001 | 25 |
| 3 | 11000 | 24 |
| 4 | 11111 | 31 |
| 5 | 11101 | 29 |
| 6 | 00000 | 0 |
| 7 | 00010 | 2 |
| 8 | 00100 | 4 |
| 9 | 01000 | 8 |
| 10 | 10000 | 16 |
| 11 | 00001 | 1 |

the gestures in CIFF DS (see table 4.4 ). CapgMyo DB-a contains contain 79840 multi-channel images for all eight gestures. CIFF DS contains 109,780 multi-channel images for the eleven gestures.

For each of the labelled dataset (CapgMyo DB-a, CIFF DS) the multi-channel images for all gestures were split into training and test data by using the odd-numbered trials of each subject for training and the remaining half for testing.

## 4.3 Procedure

### 4.3.1 RA1: Sequential CNN Implementation

To evaluate the effectiveness of the sequential-CNN in pattern recognition of sEMG signals for gestures, a similar CNN architecture found in the study by W. Geng et al. was recreated using Tensorflow, the Google machine learning library [5]. This sequential-CNN is referred to as SGR-CNN (Sequential Gesture Recognition) for the remainder of the document. The difference between the CNN developed by W. Geng et al and SGR-CNN is that SGR-CNN used a four channel image as the input verse the three channel input by W. Geng et al. The Tensorflow machine learning library was chosen based on the extensive support content as well as TensorBoard which allows for the visualisations of the network parameters. The structure of the CNN developed in [5] is shown in Figure 4.4.

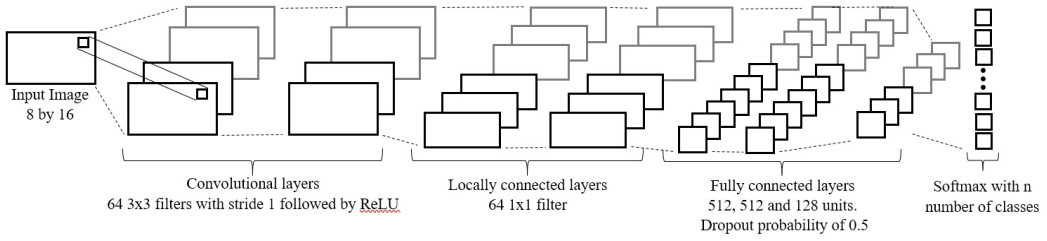



Figure 4.4: Convolutional Neural Network Developed in [5]

The 16 by 8 sEMG image as described in equation 4.1, is processed through the two convolutional layers of depth 64 and filters 3x3 with a stride of 1. No pooling layers were used. The output of the two convolutional layers is then fed into two locally connected layers of depth 64 with a filter size of 1. Three fully connected layers of size 512, 512 and 128 units follows, each with probability dropout of 0.5. The input to the first layer and output of each layer has batch normalisation. Each output of each layer utilities a ReLU activation function. Finally, the output layer is a vector of size $n$ with softmax. The output vector size $n$ corresponds to the number of gestures contained in the training and testing sets.

### 4.3.2 RA1: Validating Sequential CNN on CapgMyo DB-a

Using SGR-CNN described in the previous subsection, the gesture recognition performance was tested on the 8 isometric and isotonic gestures found in CapgMyo DB-a to validate the results found in the study by W. Geng et al. [5]. The network was trained using RMSprop with learning rate of 0.001 and L2 regularization weight decay of 0.001. SGR-CNN was trained for 18 epochs, for the first 12 epochs a batch size of 998 images was used, the last 6 epochs used a batch size of 2495 images. The expectation was that the average accuracy of gesture recognition across all subjects should be comparable with the 86.4 % obtained for the CapgMyo DB-a with power-line interference removed achieved by Geng et al. in[5].

### 4.3.3 RA2: Test Sequential Classification on CIFF DS

The sequential recognition accuracy of the SGR-CNN on the 11, individual and combined finger flexions found in the CIFF DS was tested using the same training methodology described in subsection 4.3.2.

Additionally, the recognition accuracy was tested as a function of the number of gestures. This was done for 5, 7 and 9 gestures selected from the CIFF DS. The gestures were selected based on the following method;

$$a_k = \frac{\mathrm{k}+1}{2}, k \in \{5, 7, 9\} \tag{4.2}$$

$$b_k = \frac{\mathrm{k}-1}{2}, k \in \{5, 7, 9\} \tag{4.3}$$

Where $k$ is the number of gestures contained in the test and training set.

$$\begin{aligned} \text{CIFF subset}_k = \mathit{combinations}(\text{CIFF DS}(1..6), a_k) \\ \cup \mathit{combinations}(\text{CIFF DS}(7..11), b_k) \end{aligned} \tag{4.4}$$

The $CIFFsubset_k$ contained 200, 150 and 30 gesture combinations for k equal to 5, 7 and 9 gestures respectively. In order to manage network training time, 25, 20 and 15 gestures sets were randomly chosen from CIFF $\text{subset}_5$, CIFF $\text{subset}_7$ and CIFF $\text{subset}_9$ respectively. See Tables 4.5, 4.6 and 4.7 for list of gestures chosen for each CIFF subset.

### 4.3.4 RA3: CNN Adaptation to Multi-Label

To support multi-label classification the output layer of the SGR-CNN was changed. The output layer changes required a fixed output vector size of five, a class for each finger flexion. Additionally, the softmax function was replaced with sigmoid activation function to allow for multi-label classification as detailed in the literature review subsection 2.5.2. This new network architecture is referred to as ML-CNN (multi-label).

In order to evaluate the gesture recognition accuracy of ML-CNN architecture, the prediction from the networks final layer was dynamically rounded to zero or one (non-flexion/flexion) based on the "best threshold" for that particular evaluation. The "best threshold" is the rounding value of the multi-label output that produces the overall best classification accuracy. The rounded prediction was then converted from a binary value to a demical value. The demical values were then compared with the demical ground truth (see Table 4.4). Figure 4.5 provides a visual representation of how the individual finger flexion classes were converted to gestures for gesture recognition evaluation.

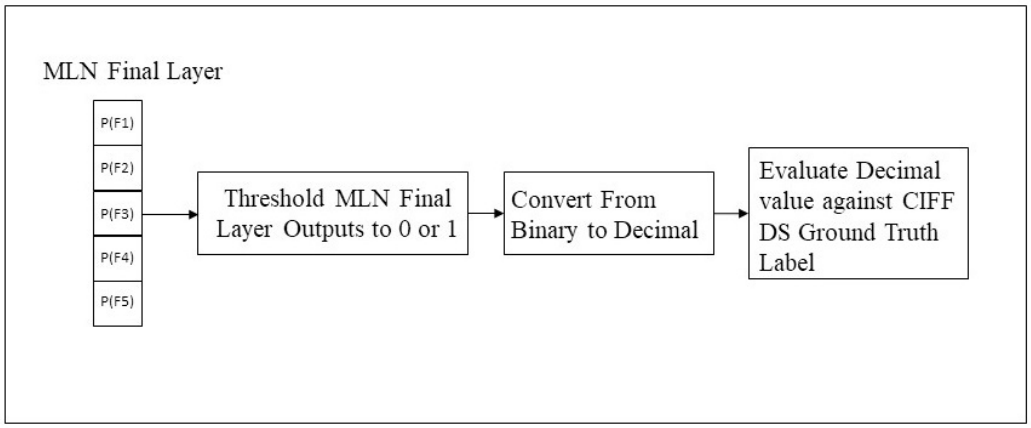



Figure 4.5: Conversion from Finger Flexion Encoding to Decimal Ground Truth

### 4.3.5 RA3: Test Multi-label Finger Flexion Classification on CIFF DS

Using ML-CNN the gesture recognition accuracy for the eleven gestures from CIFF DS was tested. Additionally, the gesture recognition accuracy as a function of the number of gestures was tested for the CIFF subsets described in 4.3.3 for 5, 7 and

9 gestures (see Tables 4.5, 4.6 and 4.7 ). The network was trained using RMSprop with a learning rate of 0.001 and L2 regularization weight decay of 0.001. ML-CNN was trained for 24 epochs, for the first 16 epochs a batch size of 998 images was used, the last 8 epochs used a batch size of 2495 images. The gesture recognition accuracy was calculated based on the method described in subsection 4.3.4.

### 4.3.6 RA4: Multi-label Finger Flexion Classification on Unseen CIFF DS gestures

Using ML-CNN the recognition accuracy for the eleven gestures from CIFF DS is tested. The network was trained using the CIFF subset$_5$ (see table 4.5 ) as well as CIFF DS gestures 7, 8, 9, 10 and 11 (a set of single finger flexions, see fig 4.1). The network was trained using RMSprop with a learning rate of 0.001 and L2 regularization weight decay of 0.001. ML-CNN was trained for 24 epochs, for the first 16 epochs a batch size of 998 images was used, the last 8 epochs used a batch size of 2495 images. The gesture recognition accuracy is calculated based on the method described in subsection 4.3.4.

Tables 4.5, 4.6 and 4.7 indicate with a “x” which of the eleven gestures are included in each CIFF subset. The CIFF subsets are references as CIFF subset $_{g-i}$ where $g$ is the number of gestures in the subset and $i$ is the CIFF subset index. Table 4.5, above contains twenty-five subsets each containing five gestures. The twenty-five randomly selected sets of five gestures equates to 12.5 % of the total number of combinations contained in CIFF subset $_5$.

Table 4.5: Five Gesture CIFF $\text{Subset}_5$

| CIFF DS Index | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Decimal Value | | 30 | 25 | 24 | 31 | 29 | 0 | 2 | 4 | 8 | 16 | 1 |
| CIFF subset index | 1 | x | x | | | x | | x | x | | | |
| | 2 | x | x | x | | | | x | | x | | |
| | 3 | x | x | x | | | | | x | | x | |
| | 4 | x | x | x | | | | | x | | | x |
| | 5 | x | x | | | | x | | x | | x | |
| | 6 | x | | | x | x | | | | x | | x |
| | 7 | x | | | | x | x | | | x | x | |
| | 8 | x | | x | x | | | x | | | | x |
| | 9 | x | | | x | | x | x | x | | | |
| | 10 | | x | | x | x | | x | | x | | |
| | 11 | | x | x | x | | | x | | | x | |
| | 12 | | x | x | x | | | x | | | | x |
| | 13 | | x | x | x | | | | x | | x | |
| | 14 | | x | x | x | | | | | x | x | |
| | 15 | | x | | x | | x | x | | | | x |
| | 16 | | x | | x | | x | | | x | | x |
| | 17 | | x | x | | | x | x | x | | | |
| | 18 | | x | x | | | x | | x | | x | |
| | 19 | | | x | x | x | | x | x | | | |
| | 20 | | | | x | x | x | x | | x | | |
| | 21 | | | | x | x | x | | x | | | x |
| | 22 | | | x | | x | x | x | | x | | |
| | 23 | | | x | | x | x | x | | | x | |
| | 24 | | | x | x | | x | x | | | x | |
| | 25 | | | x | x | | x | x | | | | x |

Table 4.6, above contains twenty sets each containing seven gestures. The twenty randomly selected sets of seven gestures equates to 13.33 % of the total number of combinations contained in CIFF subset $_7$.

Table 4.6: Seven Gesture CIFF Subset$_7$

| CIFF DS Index | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Decimal Value | | 30 | 25 | 24 | 31 | 29 | 0 | 2 | 4 | 8 | 16 | 1 |
| CIFF subset index | 1 | x | x | | x | x | | x | x | x | | |
| | 2 | x | x | x | | x | | x | | | x | x |
| | 3 | x | x | | | x | x | x | | x | | x |
| | 4 | x | x | | | x | x | | x | | x | x |
| | 5 | x | x | | x | | x | | x | x | | x |
| | 6 | x | x | x | | | x | x | | x | | x |
| | 7 | x | | x | x | x | | x | x | x | | |
| | 8 | x | | x | x | x | | | x | x | x | |
| | 9 | x | | | x | x | x | | x | | x | x |
| | 10 | x | | x | | x | x | | x | x | x | |
| | 11 | x | | x | | x | x | | | x | x | x |
| | 12 | x | | x | x | | x | | x | x | x | |
| | 13 | | x | x | x | x | | | | x | x | x |
| | 14 | | x | | x | x | x | | x | x | | x |
| | 15 | | x | x | | x | x | x | | | x | x |
| | 16 | | x | x | | x | x | | x | x | x | |
| | 17 | | | x | x | x | x | x | x | x | | |
| | 18 | | | x | x | x | x | x | x | | | x |
| | 19 | | | x | x | x | x | x | | x | x | |
| | 20 | | | x | x | x | x | | x | x | | x |

Table 4.7, above contains fifteen sets each containing seven gestures. The fifteen randomly selected sets of nine gestures equates to 50% of the total number of combinations contained in CIFF subset $_9$.

Table 4.7: Nine Gesture CIFF $\text{Subset}_9$

| CIFF DS Index | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Decimal Value | | 30 | 25 | 24 | 31 | 29 | 0 | 2 | 4 | 8 | 16 | 1 |
| CIFF subset index | 1 | x | x | x | x | x | | x | x | x | x | |
| | 2 | x | x | x | x | x | | x | x | | x | x |
| | 3 | x | x | x | x | x | | x | | x | x | x |
| | 4 | x | x | | x | x | x | x | x | | x | x |
| | 5 | x | x | | x | x | x | x | | x | x | x |
| | 6 | x | x | | x | x | x | | x | x | x | x |
| | 7 | x | x | x | | x | x | x | x | x | | x |
| | 8 | x | x | x | | x | x | x | x | | x | x |
| | 9 | x | x | x | x | | x | x | x | x | x | |
| | 10 | x | x | x | x | | x | x | | x | x | x |
| | 11 | x | x | x | x | | x | | x | x | x | x |
| | 12 | x | | x | x | x | x | x | x | | x | x |
| | 13 | x | | x | x | x | x | x | | x | x | x |
| | 14 | x | | x | x | x | x | | x | x | x | x |
| | 15 | | x | x | x | x | x | x | | x | x | x |

# Chapter 5

# Results and Discussion

In this chapter, the results for each of the research areas will be presented. The results for each research area will be discussed by examining the main observations and exceptions. The results for RA1, RA2, RA3 and RA4 are examined in Sections 5.1 through 5.4. In Section 5.1 (RA1) the sequential CNN trained and tested on HD-sEMG data is validated as a means of gesture classification by matching the results obtained in the study by W. Geng [5]. In Section 5.2 (RA2) the SGR-CNN gesture recognition mean accuracy is shown to decrease as the number of gestures increase, as discussed in the literature (see 2.5) [3]. The gesture recognition accuracy of the SGR-CNN will form the baseline to which the ML-CNN performance will be compared. In section 5.3 (RA3) results of the tests on the ML-CNN are presented. It is shown in section 5.4 that the ML-CNN can recognize gestures unseen during training by recognizing the individual finger flexions that the gestures are comprised of. Finally, in section 5.5 the research questions are answered.

## 5.1 RA 1: Validate the classification performance of Sequential CNN on HD-sEMG Images

In order to validate the classification performance of a sequential CNN on HD-sEMG data, the network architecture described in Section 4.3.1 was trained and tested as prescribed in subsection 4.3.2. Figure 5.1 shows the classification accuracy, the mean precision and the mean recall for each of the subjects in CapgMyo DBa. The mean accuracy across all the subjects was 86.44% which matches and validates the results found by Geng et al in [5]. This result confirms that the CNN when trained on HD-sEMG, provides state of the art gesture recognition performance.

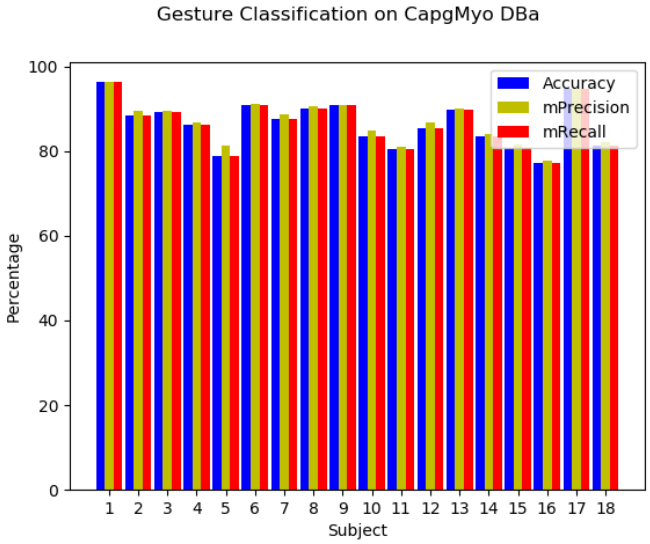
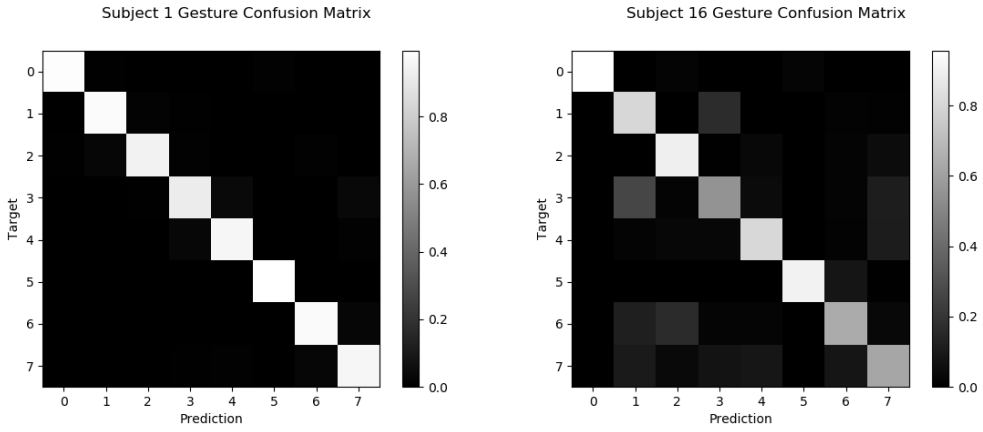



Figure 5.1: Sequential Gesture Recognition Accuracy on CapgMyo DBa

The subject who's HD-sEMG data had the best classification performance was subject 1 with 96.35% accuracy, while subject 16 had the lowest classification accuracy of 77.35%. The confusion matrix for subjects 1 and 16 are illustrated in fig 5.2.




(a) (b)

Figure 5.2: Confusion Matrix Results for Subjects 1 and 16 from RA1

Inspecting the confusion matrix for subject 16 shows there is gesture misrecognition between gesture 1 and gestures 3. Gestures 1 and 3 refer to CapgMyo DBa gestures 2 and 4, as seen in fig 5.3 [5]. The two CapgMyo DBa gestures share three finger movements, flexion of the thumb and the extension of the index and middle finger. It is possible that these similarities between the two CapgMyo DBa gestures is more pronounced for subject 16 and causes the gesture misclassification.

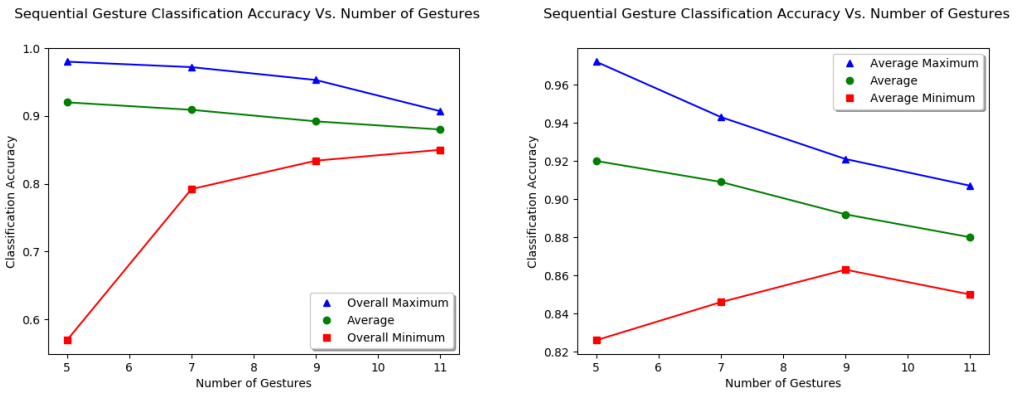


(a) Gesture 2 (b) Gesture 4

Figure 5.3: CapgMyo DBa Gestures 2 and 4 [5]

## 5.2 RA2: Sequential Gesture Recognition on CIFF DS

As described in Section 4.3.3 the sequential gesture classification accuracy was tested for CIFF DS gesture subsets containing 5, 7, 9 and 11 gestures to measure classification as a function of the number of gestures. These results additionally created a baseline to which multi-label classification could be compared. Figure 5.4 shows two plots, fig 5.4a is the plot of overall maximum, overall minimum and average of each of the CIFF DS gesture subsets. Figure 5.4b is a plot of average maximum, average minimum and average of each of the CIFF DS gesture subsets.




(a) (b)

Figure 5.4: Sequential Gesture Recognition Accuracy as a Function of the Number of Gestures

Figure 5.4 shows that the eleven gesture average accuracy across all the subjects was 88.0%. The results of RA2 showed that as the number of gestures increased the classification accuracy dropped. This was expected, as discussed in the literature review, as the number of classes increase the more difficult it becomes to recognise

the differences between the gestures sEMG patterns.

The overall minimum accuracy seen in fig 5.4 for a five gesture set, was subject 3, CIFF $\text{subset}_{5-21}$. The overall minimum accuracy was found to be 56.9 %, with a mean precision of 54.3% and mean recall of 56.9%. The confusion matrix for subject 3 can be seen in fig 5.5.

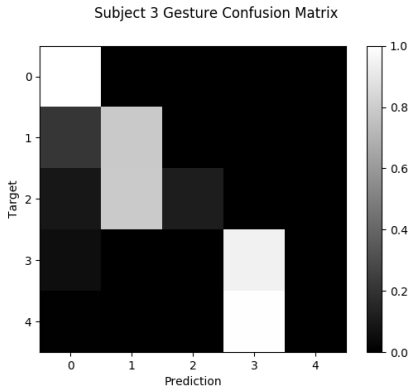


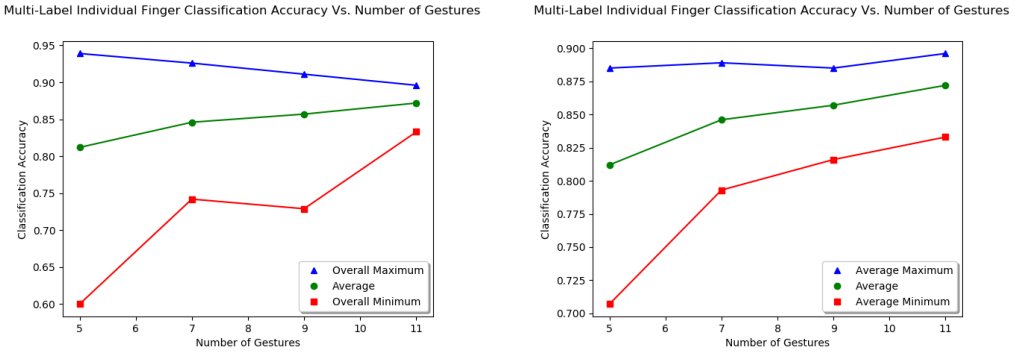

Figure 5.5: Confusion Matrix Results for Subject 3 on CIFF $\text{Subset}_{5-21}$

Inspecting the gestures contained in CIFF $\text{subset}_{5-21}$ and the confusion matrix for this experiment, there is severe confusion between gestures CIFF DS gestures 5 and 6, and, 8 and 11. Considering the fact that CIFF DS gestures 5 and 6 share four of the same finger flexions and CIFF DS gestures 8 and 11 share 3 three non-flexions, this likely explains the confusion. The SGR-CNN was not able to learn the patterns that separate theses similar gestures. It was also interesting to note of that variation between best and worst gesture classification decreased as the number of gestures increased. A possible explanation for this is, as the number of gestures increased so did the number of training images, therefore providing more information for the CNN to learn the sEMG patterns. This could be tested by limiting the training images in each of CIFF $\text{subsets}_{7,9}$ and CIFF DS to the same number of images found in each of the CIFF $\text{subset}_5$.

## 5.3 RA3: Multi-Label Gesture Recognition on CIFF DS

As described in subsection 4.3.5 the multi-label gesture classification accuracy was tested for CIFF $\text{subsets}_{5,7,9}$ and CIFF DS containing 5, 7, 9 and 11 gestures. This

was done to measure the gesture recognition accuracy as a function of the number of gestures. Figure 5.6 shows two plots, fig 5.6a is the plot of overall maximum, minimum and average accuracy taken for each of the CIFF subsets$_{5,7,9}$ and CIFF DS. Figure 5.4b is a plot of average maximum, minimum and average accuracy taken for each of the CIFF gesture sets.




(a) (b)

Figure 5.6: Multi-Label Gesture Recognition Accuracy as a Function of the Number of Gestures

As seen in fig 5.6 as the number of gestures contained in each CIFF subset increase so does the recognition average accuracy across each CIFF subset. It must be noted that the increase in average accuracy does seem to provide a diminishing return as the number of gestures contained in each subset increases, hence it is unlikely to be a linear relationship.

## 5.4 RA4: Multi-Label Gesture Recognition on Unseen CIFF DS Gestures

In order to test if the patterns learned by the ML-CNN to decode individual finger flexions from gestures were generalised across different gestures, the gestures contained in CIFF subset$_5$ were used to train the ML-CNN as described in subsection 4.3.6. The gesture recognition accuracy was then tested on the entire CIFF DS. The results of the tests have been divided into two parts; The set of five gestures that resulted in best gesture recognition accuracy. The second part is the gesture recognition results for CIFF subset composed of CIFF gesture indexes, 7, 8, 9, 10 and 11 (a set of single finger flexions, see fig 4.1 ).

### 5.4.1 Eleven Gesture Recognition Performance for Best Five Gesture Subset

The five gesture set that resulted in the best performance was CIFF $\text{subset}_{5-1}$ . The mean accuracy across all subject was 81.7% as seen in Table 5.1. The subject whose gestures were best recognized for CIFF $\text{subset}_{5-1}$ was subject 8, with a recognition accuracy of 85.3%. The worst recognition performance for CIFF $\text{subset}_{5-1}$ was subject 1, with a recognition accuracy of 77.6%. The precision, recall and f1-score

Table 5.1: Gesture Recognition Performance Summary for CIFF $\text{Subset}_{5-1}$

| Subject | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| Accuracy ( %) | 77.6 | 83.5 | 80.8 | 82.3 | 81.8 | 80.5 | 81.5 | 85.3 |
| Mean Accuracy (%) | 81.7 | | | | | | | |

for subjects 1 and 8 can be seen in Tables 5.2 and 5.3 respectively. The grey rows indicate gestures which have been used for training and testing, while the white rows indicate gestures only used for testing.

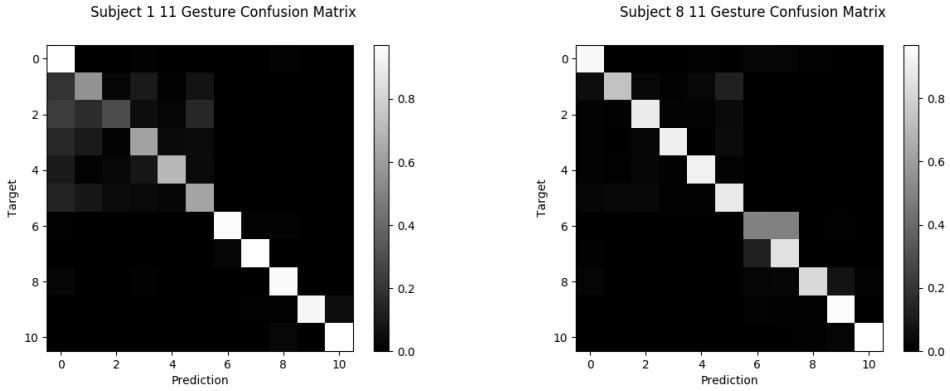
Table 5.2: Subject 1 Gesture Recognition Performance Report for CIFF $\text{Subset}_{5-1}$

| CIFF DS Index | Ground Truth Decimal Value | Precision | Recall | f1-score |
|---|---|---|---|---|
| 6 | 0 | 0.534 | 0.970 | 0.689 |
| 11 | 1 | 0.600 | 0.557 | 0.577 |
| 7 | 2 | 0.729 | 0.292 | 0.417 |
| 8 | 4 | 0.684 | 0.617 | 0.649 |
| 9 | 8 | 0.883 | 0.698 | 0.780 |
| 10 | 16 | 0.671 | 0.623 | 0.646 |
| 3 | 24 | 0.966 | 0.956 | 0.961 |
| 2 | 25 | 0.976 | 0.969 | 0.972 |
| 5 | 29 | 0.940 | 0.953 | 0.946 |
| 1 | 30 | 0.997 | 0.934 | 0.965 |
| 4 | 31 | 0.947 | 0.972 | 0.960 |

The confusion matrix for subjects 1 and 8 are illustrated in fig 5.7. Figure 5.7a illustrates the confusion matrix for subject 1. The confusion matrix for subject 1 shows that CIFF gestures 11, 7 and 8 had the worst misclassification. The confusion matrix for subject 8 is illustrated in fig 5.7b. Subject 8's confusion matrix shows

Table 5.3: Subject 8 Gesture Recognition Performance Report for CIFF $\text{Subset}_{5-1}$

| CIFF DS Index | Ground Truth Decimal Value | Precision | Recall | f1-score |
|---|---|---|---|---|
| 6 | 0 | 0.861 | 0.946 | 0.901 |
| 11 | 1 | 0.938 | 0.738 | 0.826 |
| 7 | 2 | 0.889 | 0.896 | 0.893 |
| 8 | 4 | 0.960 | 0.907 | 0.933 |
| 9 | 8 | 0.943 | 0.916 | 0.929 |
| 10 | 16 | 0.801 | 0.885 | 0.841 |
| 3 | 24 | 0.730 | 0.495 | 0.590 |
| 2 | 25 | 0.613 | 0.855 | 0.714 |
| 5 | 29 | 0.969 | 0.824 | 0.890 |
| 1 | 30 | 0.904 | 0.954 | 0.928 |
| 4 | 31 | 0.985 | 0.969 | 0.977 |




(a) (b)

Figure 5.7: Multi-Label Gesture Recognition Confusion Matrix

CIFF gesture 3, the flexion of the ring and little finger, is confused with CIFF gesture 10, the flexion of the little finger. It is common for people to struggle to perform an isolated little finger flexion, this is what likely caused the misclassification of CIFF gesture 3 in subject 8.

### 5.4.2 Eleven Gesture Performance for Five Single Individual Finger Flexions

In order to gain insight into how individual finger flexions may combine to form a gesture, the ML-CNN was trained using only individual finger flexions. As seen in

Table 5.4, the mean accuracy across all subject was 81.1%, 0.6% lower than that found for the best combinations of gestures in CIFF subset$_5$ (see Table 5.1). Subject 2 had the best gesture recognition accuracy across all eleven gestures while subject 1 again had the worst accuracy. The precision, recall and f1-score for subjects 1 and 2 can be seen in Tables 5.5 and 5.6 respectively. The grey rows indicate gestures which have been used for training and testing, the white rows indicate gestures which were used for testing.

Table 5.4: Gesture Recognition Performance Summary for CIFF Subset$_{5-1}$

| Subject | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| Accuracy ( %) | 0.752 | 0.855 | 0.758 | 0.837 | 0.802 | 0.828 | 0.814 | 0.842 |
| Mean Accuracy (%) | 81.1 | | | | | | | |

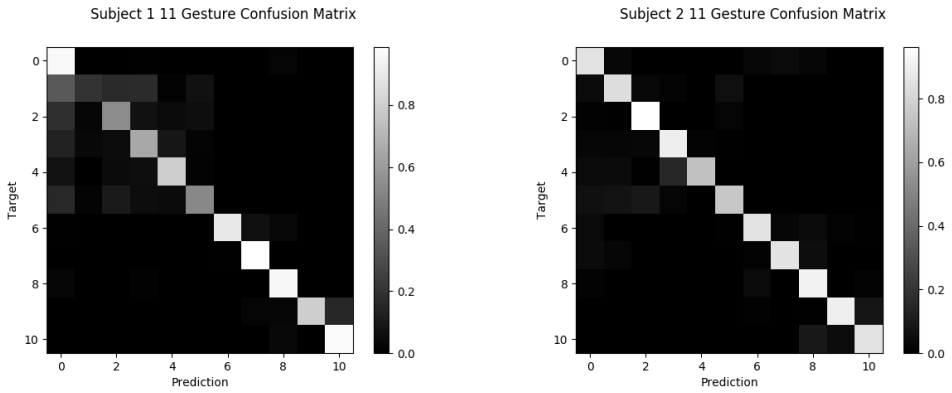
Table 5.5: Subject 1 Gesture Recognition Performance Report for Single Finger Flexion Gesture Set

| CIFF DS Index | Ground Truth Decimal Value | Precision | Recall | f1-score |
|---|---|---|---|---|
| 6 | 0 | 0.511 | 0.961 | 0.667 |
| 11 | 1 | 0.722 | 0.200 | 0.313 |
| 7 | 2 | 0.600 | 0.541 | 0.569 |
| 8 | 4 | 0.642 | 0.650 | 0.646 |
| 9 | 8 | 0.810 | 0.794 | 0.802 |
| 10 | 16 | 0.775 | 0.525 | 0.626 |
| 3 | 24 | 0.987 | 0.899 | 0.941 |
| 2 | 25 | 0.924 | 0.988 | 0.955 |
| 5 | 29 | 0.892 | 0.956 | 0.923 |
| 1 | 30 | 1.000 | 0.796 | 0.887 |
| 4 | 31 | 0.863 | 0.966 | 0.911 |

It is of interest to note that CIFF DS gestures 7-11 came from a single session of Capgmyo DB-c, the remainder of the gestures contained in CIFF DS were recorded during a different session. The gesture recognition results for subject 2 show that the mean precision for CIFF DS gesture 1-6 was 87%. This result indicates that individual finger flexion sEMG patterns are generalised across gestures and sessions. The confusion matrix for subjects 1 and 2 are illustrated in fig 5.8

Table 5.6: Subject 2 Gesture Recognition Performance Report for Single Finger Flexion Gesture Set

| CIFF DS Index | Ground Truth Decimal Value | Precision | Recall | f1-score |
|---|---|---|---|---|
| 6 | 0 | 0.755 | 0.856 | 0.802 |
| 11 | 1 | 0.808 | 0.832 | 0.820 |
| 7 | 2 | 0.859 | 0.963 | 0.908 |
| 8 | 4 | 0.825 | 0.896 | 0.859 |
| 9 | 8 | 0.972 | 0.727 | 0.832 |
| 10 | 16 | 0.889 | 0.750 | 0.813 |
| 3 | 24 | 0.903 | 0.857 | 0.880 |
| 2 | 25 | 0.922 | 0.858 | 0.889 |
| 5 | 29 | 0.805 | 0.912 | 0.855 |
| 1 | 30 | 0.936 | 0.900 | 0.918 |
| 4 | 31 | 0.899 | 0.854 | 0.876 |




(a) (b)

Figure 5.8: Multi-Label Gesture Recognition Confusion Matrix

Figure 5.8a and 5.8b illustrate the confusion matrix for subject 1 and 2 respectively. Subject 1 again showed confusion across gestures 11, 7 and 8.

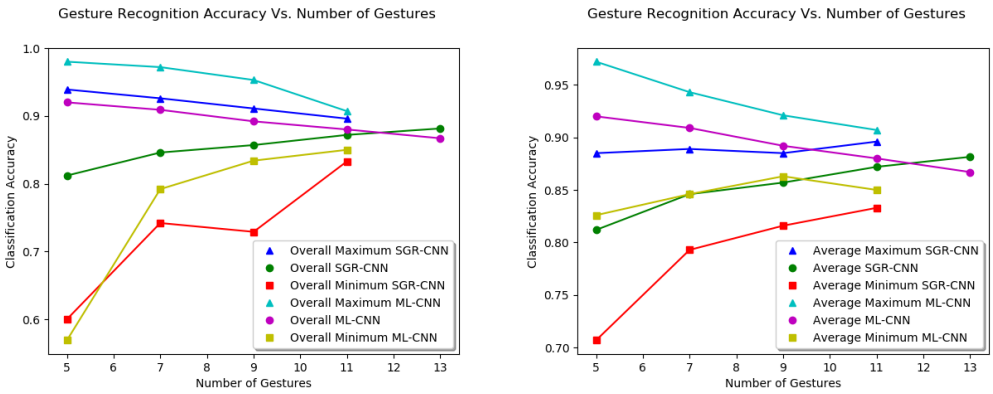
## 5.5 Answering the Research Questions

In the following subsections research questions 1 and 2 will be answered and discussed.

### 5.5.1 RQ 1

The first of the research questions asked in section 3.1, can gesture recognition accuracy be improved from the current state of the CNN gesture recognition method by using a multi-label CNN to recognize five individual finger flexions and infer the correct gesture?
In section 5.1 the SGR-CNN method for gesture recognition based on the paper by W. Geng et al. was shown to match the gesture recognition accuracy achieved W. Geng et al. [5].
In section 5.2 the SGR-CNN was tested using gesture sets containing five, seven, nine and eleven gestures. It was shown that SGR-CNN gesture recognition accuracy decreases as the number of gestures increase. The results for SGR-CNN gesture recognition tests created the baseline for comparison with the ML-CNN. In section 5.3 the ML-CNN was tested using gesture sets containing five, seven, nine and eleven gestures. It was shown that ML-CNN gesture recognition accuracy **increases as** the number of gestures increase. This demonstrates a major triumph for the ML-CNN over SGR-CNN.

Comparing the results obtained for RA 2 and RA 3 allow us to answer RQ 1: In fig 5.9 the gesture recognition accuracy for SGR-CNN and ML-CNN are plotted. Figure 5.9 shows that for all the CIFF DS subsets containing five, seven, nine and eleven gestures the average recognition accuracy of each subset is higher for the SGR-CNN than ML-CNN. It can be concluded for gesture sets containing eleven gestures or less, the ML-CNN does not improve gesture classification. It should be noted however that the two networks (SGR and ML) recognition performance does seem to be converging at eleven gestures. Moreover, by extrapolating the average gesture recognition for a set containing twelve gestures for SGR and ML-CNN, the ML-CNN may provide better recognition performance. Due to the limits of this study, a twelfth and thirteenth gesture containing unique individual finger flexions could not be tested.




(a) (b)

Figure 5.9: Comparison of SGR and ML-CNN Gesture Recognition Accuracy

### 5.5.2 RQ 2

Research Question 2 asked in section 3.1, using a 128 channel sEMG dataset, to what accuracy can eleven gestures be recognized by the multi-label CNN trained on only five gestures?

In section 5.4 the ML-CNN was trained using gesture sets containing five gestures, CIFF subset$_5$ and tested on a gesture set containing eleven gestures, CIFF DS. Examining the results obtained in RA4 provided that answer to RQ 2. It was shown that an 81.7% gesture recognition accuracy could be achieved across all subjects when the ML-CNN was trained on CIFF subset$_{5-1}$. The eleven gesture recognition accuracy of ML-CNN, when trained on CIFF subset$_{5-1}$ was therefore 6.3% less accurate than SGR-CNN and 5.5% less accurate than ML-CNN train on eleven gestures. Additionally, it was shown that ML-CNN could be trained on individual finger flexions only and still achieve a gesture recognition accuracy of 81.1% across all subjects. The results for RA 4 showed that individual finger flexion patterns are generalised across at least the eleven gestures contained in the CIFF DS.

## 5.6 Summary

The results for each of the investigated research areas were presented in this chapter. The main points of interest from each research were discussed and the relevant figures illustrated. The success of the multi-label convolutional neural network in all research areas is explored and the significance considered. The key findings from each of the research areas were used to answer Research Questions 1 and 2. The chapter

to follow concludes the research by discussing the main findings and suggesting recommendations for future work.

# Chapter 6

# Conclusion

The conclusion is separated into three sections, section 6.1 provides a summary of the study. Section 6.2 summaries the main findings of the study. The final section 6.3 makes recommendations for future work that can lead on from the developments made in this study.

## 6.1 Summary

Improvements in machine learning have enabled the development of systems that achieved high accuracy in gesture recognition from surface electromyography (sEMG), however, these systems have only ever provided a limited number of gestures. As shown in this study and others, the gesture recognition accuracy of the sEMG signals decreases as the number of gestures classes increase. The use of a sequential gesture recognition orientated control system for prostheses additionally places a burden on the user to remember all the gestures afforded to them.

This study consisted of two aims, the first was to improve gesture recognition by using a multi-label convolutional neural network (CNN) to decode five classes, one for each finger flexion and from the recognition of these finger flexions to infer the correct gesture. The second was to determine if the sEMG patterns specific to individual finger flexions were generalised across different hand gestures. Determining how sEMG signals for individual finger movements combine to form a gesture may allow for a control system which is focused toward decoding the extent of finger involvement rather than classifying hand gestures as a whole. A system based on decoded individual finger actions would reduce the total class count and possibly provide an sEMG analysis technique which would move toward more dexterous

prosthesis.

In this study it is shown that a multi-label CNN used to decode individual finger flexions does not provide better gesture recognition accuracy for eleven gesture or less, However unlike conventional gesture recognition approaches, the recognition accuracy **does not decrease** with an increasing number of gesture classes. This result indicates that for larger gesture sets (twelve of more) a multi-label CNN used to decode individual finger flexions may improve gesture recognition. Additionally, it is shown that the sEMG patterns specific to individual finger flexions are generalised across different gestures. This is demonstrated by training the multi-label CNN on a gesture set comprised of only the five individual finger flexions and testing the gesture recognition accuracy across an additional six hand gestures.

## 6.2 Main Findings

The main finding of this study are listed below, in order of appearance:

1. The sequential gesture recognition accuracy of 86.4 % achieved by W.Geng et al was matched with the sequential gesture recognition CNN developed in this study.

2. The gesture recognition accuracy for the multi-label CNN used to decode five individual finger flexions to infer hand gestures was less than that found for the sequential gesture recognition CNN on the CIFF dataset.

3. The multi-label CNN used to decode five individual finger flexions to infer hand gestures was shown to have **increasing** gesture recognition accuracy as the number of gestures in the training sets increased, this is a major improvement from sequential gesture recognition CNN, which was shown to have **decreasing** accuracy for the same test.

4. The multi-label CNN used to decode five individual finger flexions showed that the sEMG patterns of the individual finger flexions are generalised across other gestures containing those same finger flexions.

## 6.3 Recommendations for Future Work

The results of this study have provided two main areas of research for future work. In this study, the number of gestures were limited to those captured in the Capgmyo database containing unique individual finger flexion combinations [5]. This meant that only eleven gestures could be tested. It was shown by extrapolation that the relationship between recognition accuracy and the number of gestures contained in a set that the multi-label CNN could provide better classification for twelve or more gestures than the sequential gesture recognition CNN. Thus the first area of research for future work would be to test if in fact the multi-label CNN approach to gesture recognition would provide better gesture recognition accuracy than a sequential gesture recognition approach over gestures sets containing twelve or more gestures.

The second area of future research could be posed as an optimisation study. In this study, the optimal five gesture set for training the multi-label CNN to recognise eleven gestures was determined from twenty-five variations of the eleven gestures contained in the CIFF dataset. This provides a number of optimisation questions; What is the fewest number of gestures that the multi-label CNN could be trained on to achieve generalised gesture recognition accuracy? Which gestures would provide the best generalised gesture recognition accuracy?

A possible extension for multi-label CNN could be in discrimination electroencephalogram (EEG) data to discern the movements of the shoulder, albow and wrist. In a similar way to how the 128 channel sEMG data was converted to sEMG images the EEG electrode channels could to be converted to EEG images.

# Bibliography

[1] L. Pan, D. Zhang, N. Jiang, X. Sheng, and X. Zhu, "Improving robustness against electrode shift of high density EMG for myoelectric control through common spatial patterns," *Journal of NeuroEngineering and Rehabilitation*, vol. 12, p. 110, dec 2015.

[2] K. Englehart and B. Hudgins, "A robust, real-time control scheme for multi-function myoelectric control," *IEEE Trans Biomed Eng*, vol. 50, pp. 848–854, jul 2003.

[3] C. W. Ke Lin, X. Huang, Q. Ding, and X. Gao, "A Robust Gesture Recognition Algorithm Based on Surface EMG," *Conference on Advanced Computational Interfaces*, pp. 131–136, 2015.

[4] F. Tenore, A. Ramos, A. Fahmy, S. Acharya, R. Etienne-Cummings, and N. Thakor, "Decoding of individuated finger movements using surface Electromyography," *Biomedical Engineering, IEEE Transactions on*, vol. 56, no. 5, pp. 1427–1434, 2009.

[5] W. Geng, Y. Du, W. Jin, W. Wei, Y. Hu, and J. Li, "Gesture recognition by instantaneous surface EMG images.," *Scientific reports*, vol. 6, no. October, p. 36571, 2016.

[6] M. B. I. Raez, M. S. Hussain, F. Mohd-Yasin, M. Reaz, M. S. Hussain, and F. Mohd-Yasin, "Techniques of EMG signal analysis: detection, processing, classification and applications.," *Biological procedures online*, vol. 8, no. 1, pp. 11–35, 2006.

[7] H. Chen, R. Tong, M. Chen, Y. Fang, and H. Liu, "A Hybrid Cnn-SVM Classifier for Hand Gesture Recognition with Surface Emg Signals," in *Proceedings - International Conference on Machine Learning and Cybernetics*, vol. 2, pp. 619–624, IEEE, jul 2018.

[8] Y. Du, W. Jin, W. Wei, Y. Hu, and W. Geng, "Surface EMG-Based Inter-Session Gesture Recognition Enhanced by Deep Domain Adaptation," *Sensors*, vol. 17, p. 458, feb 2017.

[9] E. Biddiss and T. Chau, "Upper limb prosthesis use and abandonment: A survey of the last 25 years," sep 2007.

[10] K. Nazarpour, C. Cipriani, D. Farina, and T. Kuiken, "Guest editorial," jul 2014.

[11] F. Cordella, A. L. Ciancio, R. Sacchetti, A. Davalli, A. G. Cutti, E. Guglielmelli, and L. Zollo, "Literature review on needs of upper limb prosthesis users," 2016.

[12] B. Peerdeman, D. Boere, H. Witteveen, R. Huis in 'tVeld, H. Hermens, S. Stramigioli, H. Rietman, P. Veltink, and S. Misra, "Myoelectric forearm prostheses: State of the art from a user-centered perspective," *The Journal of Rehabilitation Research and Development*, vol. 48, no. 6, p. 719, 2011.

[13] T. Feix, J. Romero, H. B. Schmiedmayer, A. M. Dollar, and D. Kragic, "The GRASP Taxonomy of Human Grasp Types," *IEEE Transactions on Human-Machine Systems*, vol. 46, no. 1, pp. 66–77, 2016.

[14] Shadow Robot, "Shadow Dexterous Hand," 2018.

[15] N. Jiang, J. L. Vest-Nielsen, S. Muceli, and D. Farina, "EMG-based simultaneous and proportional estimation of wrist/hand dynamics in uni-Lateral trans-radial amputees," *Journal of NeuroEngineering and Rehabilitation*, vol. 9, no. 1, p. 42, 2012.

[16] A. Gallina and A. Botter, "Spatial localization of electromyographic amplitude distributions associated to the activation of dorsal forearm muscles," *Frontiers in Physiology*, vol. 4 DEC, p. 367, dec 2013.

[17] Y. Fang and H. Liu, "Robust sEMG electrodes configuration for pattern recognition based prosthesis control," in *Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics*, vol. 2014-Janua, pp. 2210–2215, IEEE, oct 2014.

[18] D. R. Merrill, J. Lockhart, P. R. Troyk, R. F. Weir, and D. L. Hankin, "Development of an Implantable Myoelectric Sensor for Advanced Prosthesis Control," *Artificial Organs*, vol. 35, pp. 249–252, mar 2011.

[19] P. Whittaker, *Opinion No. 20: Ethical aspects of ICT implants in the human body.* Office for Official Publications of the European Communities, 2005.

[20] J. C. Meadows, "Observations on muscle pain in man, with particular reference to pain during needle electromyography.," *Journal of neurology, neurosurgery, and psychiatry*, vol. 33, pp. 519–523, aug 1970.

[21] N. Jiang, S. Dosen, K. R. Muller, and D. Farina, "Myoelectric Control of Artificial Limbs: Is There a Need to Change Focus? [In the Spotlight]," *IEEE Signal Processing Magazine*, vol. 29, pp. 150–152, sep 2012.

[22] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, vol. 2017-Janua, pp. 6517–6525, 2017.

[23] A. Allen, "OK Google: Find the Humpback Whales — NOAA Fisheries," 2018.

[24] Y. Zhao, S. Wu, L. Reynolds, and S. Azenkot, "A Face Recognition Application for People with Visual Impairments: Understanding Use Beyond the Lab," 2018.

[25] W. Wei, Y. Wong, Y. Du, Y. Hu, M. Kankanhalli, and W. Geng, "A multi-stream convolutional neural network for sEMG-based gesture recognition in muscle-computer interface," *Pattern Recognition Letters*, dec 2017.

[26] A. D. Roche, H. Rehbaum, D. Farina, and O. C. Aszmann, "Prosthetic Myoelectric Control Strategies: A Clinical Perspective," *Current Surgery Reports*, vol. 2, p. 44, mar 2014.

[27] T. bionics, "i-limb quantum," 2018.

[28] M. Asghari Oskoei and H. Hu, "Myoelectric control systems-A survey," oct 2007.

[29] I. Vujaklija, D. Farina, and O. Aszmann, "New developments in prosthetic arm systems," *Orthopedic Research and Reviews*, vol. Volume 8, no. July, pp. 31–39, 2016.

[30] X. Chen, X. Zhang, Z.-Y. Zhao, J.-H. Yang, V. Lantz, and K.-Q. Wang, "Multiple Hand Gesture Recognition Based on Surface EMG Signal," *2007 1st International Conference on Bioinformatics and Biomedical Engineering*, pp. 506–509, 2007.

[31] G. Naik, D. Kumar, and M. Palaniswami, "Multi run ICA and surface EMG based signal processing system for recognizing hand gestures," *8th IEEE International Conference on Computer and Information Technology, 2008*, pp. 700–705, 2008.

[32] Q. Li and B. Li, "Online Finger Gesture Recognition Using Surface Electromyography Signals *," *Journal of Signal and Information Processing*, vol. 4, pp. 101–105, may 2013.

[33] T. A. Kuiken, G. Li, B. A. Lock, R. D. Lipschutz, L. A. Miller, K. A. Stubblefield, and K. B. Englehart, "Targeted muscle reinnervation for real-time myoelectric control of multifunction artificial arms," *JAMA - Journal of the American Medical Association*, vol. 301, pp. 619–628, feb 2009.

[34] G. Madhavan, "Electromyography: Physiology, Engineering and Non-Invasive Applications," *Annals of Biomedical Engineering*, vol. 33, no. 11, pp. 1671–1671, 2005.

[35] A. J. Young, L. H. Smith, E. J. Rouse, and L. J. Hargrove, "Classification of simultaneous movements using surface EMG pattern recognition," *IEEE Transactions on Biomedical Engineering*, vol. 60, pp. 1250–1258, may 2013.

[36] P. Shenoy, K. J. Miller, B. Crawford, and R. P. Rao, "Online electromyographic control of a robotic prosthesis," *IEEE Transactions on Biomedical Engineering*, vol. 55, pp. 1128–1135, mar 2008.

[37] E. J. Scheme and K. Englehart, "Electromyogram pattern recognition for control of powered upper-limb prostheses: State of the art and challenges for clinical use," *Journal of Rehabilitation Research and Development*, vol. 48, no. 6, pp. 643–659, 2011.

[38] J. Kilby, K. Prasad, and G. Mawston, "Multi-channel surface electromyography electrodes: A review," 2016.

[39] A. Stango, F. Negro, and D. Farina, "Spatial Correlation of High Density EMG Signals Provides Features Robust to Electrode Number and Shift in Pattern Recognition for Myocontrol," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 23, pp. 189–198, mar 2015.

[40] R. N. Khushaba, S. Kodagoda, M. Takruri, and G. Dissanayake, "Toward improved control of prosthetic fingers using surface electromyogram (EMG) signals," *Expert Systems with Applications*, vol. 39, pp. 10731–10738, sep 2012.

[41] E. Criswell and J. R. Cram, *Cram's introduction to surface electromyography*. Sudbury, Massachusetts: Jones and Bartlett Publishers, second edi ed., 2011.

[42] H. J. Hermens, B. Freriks, C. Disselhorst-Klug, and G. Rau, "Development of recommendations for SEMG sensors and sensor placement procedures," *Journal of Electromyography and Kinesiology*, vol. 10, pp. 361–374, oct 2000.

[43] Y. Oda, T. Sato, I. Nambu, Y. Wada, Y. Oda, T. Sato, I. Nambu, and Y. Wada, "Real-Time Reduction of Task-Related Scalp-Hemodynamics Artifact in Functional Near-Infrared Spectroscopy with Sliding-Window Analysis," *Applied Sciences*, vol. 8, p. 149, jan 2018.

[44] K. Østlie, I. M. Lesjø, R. J. Franklin, B. Garfelt, O. H. Skjeldal, and P. Magnus, "Prosthesis rejection in acquired major upper-limb amputees: A population-based survey," *Disability and Rehabilitation: Assistive Technology*, vol. 7, pp. 294–303, jul 2012.

[45] L. H. Smith, L. J. Hargrove, B. A. Lock, and T. A. Kuiken, "Determining the optimal window length for pattern recognition-based myoelectric control: Balancing the competing effects of classification error and controller delay," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 19, pp. 186–192, apr 2011.

[46] Z. Ding, C. Yang, Z. Tian, C. Yi, Y. Fu, and F. Jiang, "sEMG-based gesture recognition with convolution neural networks," *Sustainability (Switzerland)*, vol. 10, p. 1865, jun 2018.

[47] K. Xing, Z. Ding, S. Jiang, X. Ma, K. Yang, C. Yang, X. Li, and F. Jiang, "Hand gesture recognition based on deep learning method," in *Proceedings - 2018 IEEE 3rd International Conference on Data Science in Cyberspace, DSC 2018*, pp. 542–546, IEEE, jun 2018.

[48] S. Boyas, O. Maïsetti, and A. Guével, "Changes in sEMG parameters among trunk and thigh muscles during a fatiguing bilateral isometric multi-joint task in trained and untrained subjects," *Journal of Electromyography and Kinesiology*, vol. 19, pp. 259–268, apr 2009.

[49] A. Goen and D. C. Tiwari, "Review of Surface Electromyogram Signals: Its Analysis and Applications," *World Academy of Science, Engineering and Technology, International Journal of Electrical, Computer, Energetic, Electronic and Communication Engineering*, vol. 7, no. 11, pp. 1429–1437, 2013.

[50] R. H. Chowdhury, M. B. I. Reaz, M. A. B. M. Ali, A. A. A. Bakar, K. Chellappan, and T. G. Chang, "Surface electromyography signal processing and classification techniques.," *Sensors (Basel, Switzerland)*, vol. 13, pp. 12431–66, sep 2013.

[51] J. J. Baker, E. Scheme, K. Englehart, D. T. Hutchinson, and B. Greger, "Continuous detection and decoding of dexterous finger flexions with implantable myoelectric sensors," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 18, pp. 424–432, aug 2010.

[52] N. Nazmi, M. A. Rahman, S.-I. Yamamoto, S. Ahmad, H. Zamzuri, and S. Mazlan, "A review of classification techniques of emg signals during isotonic and isometric contractions," *Sensors*, vol. 16, p. 1304, 8 2016.

[53] A. Phinyomark, R. N. Khushaba, and E. Scheme, "Feature extraction and selection for myoelectric control based on wearable emg sensors," *Sensors*, vol. 18, p. 1615, 5 2018.

[54] F. Riillo, L. R. Quitadamo, F. Cavrini, E. Gruppioni, C. A. Pinto, N. C. Past, L. Sbernini, L. Albero, and G. Saggio, "Optimization of emg-based hand gesture recognition: Supervised vs. unsupervised data preprocessing on healthy subjects and transradial amputees," *Biomedical Signal Processing and Control*, vol. 14, pp. 117–125, 11 2014.

[55] W. Caesarendra, S. U. Lekson, K. A. Mustaqim, A. R. Winoto, and A. Widyotriatmo, "A classification method of hand EMG signals based on principal component analysis and artificial neural network," in *2016 International Conference on Instrumentation, Control and Automation (ICA)*, no. 3, pp. 22–27, 2016.

[56] Y. LeCun and Others, "LeNet-5, convolutional neural networks," *URL: http://yann. lecun. com/exdb/lenet*, vol. 20, 2015.

[57] U. Côtéallard, F. Nougarou, C. L. Fall, P. Gigu'ere, C. Gosselin, F. Laviolette, and B. Gosselin, "A Convolutional Neural Network for robotic arm guidance using sEMG based frequency-features," in *IEEE International Conference on Intelligent Robots and Systems*, vol. 2016-Novem, pp. 2464–2470, IEEE, oct 2016.

[58] M. Atzori, A. Gijsberts, C. Castellini, B. Caputo, A. G. M. Hager, S. Elsig, G. Giatsidis, F. Bassetto, and H. Müller, "Electromyography data for non-invasive naturally-controlled robotic hand prostheses," *Scientific Data*, vol. 1, p. 140053, dec 2014.

[59] J. Wang, Y. Yang, J. Mao, Z. Huang, C. Huang, and W. Xu, "CNN-RNN: A Unified Framework for Multi-label Image Classification," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2285–2294, 2016.

[60] W. Buntine, M. Grobelnik, D. Mladenic, J. Shawe-Taylor, J. Read, B. Pfahringer, G. Holmes, and E. Frank, "Classifier Chains for Multi-label Classification - Machine Learning and Knowledge Discovery in Databases - Lecture Notes in Computer Science," vol. 5782, pp. 254–269–269, Springer, Berlin, Heidelberg, 2009.

[61] J. Liu, W.-C. Chang, Y. Wu, and Y. Yang, "Deep Learning for Extreme Multi-label Text Classification," in *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval - SIGIR '17*, vol. 10, pp. 115–124, 2017.

[62] R. N. Khushaba, A. H. Al-Timemy, A. Al-Ani, and A. Al-Jumaily, "A Framework of Temporal-Spatial Descriptors-Based Feature Extraction for Improved Myoelectric Pattern Recognition," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, pp. 1821–1831, oct 2017.

# Appendix A

# Sequential Gesture Recognition Results

Table A.1: Complete Gesture Recognition Accuracy Results for CIFF Subset$_9$

| Subject | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| CIFF DS$_9$ | 0.873 | 0.917 | 0.880 | 0.894 | 0.872 | 0.875 | 0.893 | 0.885 |
| | 0.834 | 0.919 | 0.867 | 0.883 | 0.881 | 0.878 | 0.888 | 0.881 |
| | 0.863 | 0.915 | 0.899 | 0.903 | 0.887 | 0.874 | 0.882 | 0.887 |
| | 0.835 | 0.902 | 0.907 | 0.901 | 0.894 | 0.887 | 0.860 | 0.938 |
| | 0.873 | 0.920 | 0.921 | 0.909 | 0.882 | 0.866 | 0.879 | 0.935 |
| | 0.881 | 0.900 | 0.916 | 0.915 | 0.874 | 0.887 | 0.870 | 0.937 |
| | 0.846 | 0.922 | 0.888 | 0.856 | 0.891 | 0.881 | 0.870 | 0.883 |
| | 0.871 | 0.922 | 0.913 | 0.880 | 0.910 | 0.869 | 0.878 | 0.902 |
| | 0.860 | 0.918 | 0.865 | 0.890 | 0.892 | 0.884 | 0.891 | 0.881 |
| | 0.877 | 0.917 | 0.908 | 0.888 | 0.927 | 0.862 | 0.891 | 0.897 |
| | 0.882 | 0.907 | 0.850 | 0.890 | 0.911 | 0.867 | 0.901 | 0.880 |
| | 0.883 | 0.906 | 0.880 | 0.872 | 0.877 | 0.911 | 0.892 | 0.894 |
| | 0.847 | 0.928 | 0.914 | 0.899 | 0.884 | 0.885 | 0.878 | 0.940 |
| | 0.880 | 0.921 | 0.904 | 0.902 | 0.891 | 0.871 | 0.904 | 0.940 |
| | 0.896 | 0.892 | 0.931 | 0.905 | 0.883 | 0.903 | 0.895 | 0.953 |
| Average: | 0.867 | 0.914 | 0.896 | 0.892 | 0.890 | 0.880 | 0.885 | 0.909 |
| Max: | 0.896 | 0.928 | 0.931 | 0.915 | 0.927 | 0.911 | 0.904 | 0.953 |
| Min: | 0.834 | 0.892 | 0.850 | 0.856 | 0.872 | 0.862 | 0.860 | 0.880 |

Table A.2: Complete Gesture Recognition Accuracy Results for CIFF Subset$_7$

| Subject | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| CIFF DS$_7$ | 0.871 | 0.902 | 0.857 | 0.877 | 0.849 | 0.858 | 0.857 | 0.870 |
| | 0.898 | 0.917 | 0.908 | 0.871 | 0.846 | 0.873 | 0.882 | 0.885 |
| | 0.897 | 0.919 | 0.932 | 0.939 | 0.886 | 0.920 | 0.894 | 0.969 |
| | 0.918 | 0.920 | 0.815 | 0.889 | 0.859 | 0.923 | 0.924 | 0.886 |
| | 0.886 | 0.937 | 0.919 | 0.922 | 0.871 | 0.921 | 0.911 | 0.958 |
| | 0.904 | 0.919 | 0.924 | 0.925 | 0.888 | 0.862 | 0.910 | 0.955 |
| | 0.871 | 0.939 | 0.936 | 0.920 | 0.882 | 0.837 | 0.925 | 0.960 |
| | 0.895 | 0.927 | 0.934 | 0.937 | 0.887 | 0.915 | 0.924 | 0.965 |
| | 0.899 | 0.927 | 0.868 | 0.913 | 0.932 | 0.850 | 0.943 | 0.891 |
| | 0.911 | 0.940 | 0.925 | 0.946 | 0.896 | 0.864 | 0.892 | 0.945 |
| | 0.889 | 0.894 | 0.919 | 0.932 | 0.872 | 0.919 | 0.900 | 0.939 |
| | 0.792 | 0.920 | 0.877 | 0.889 | 0.894 | 0.888 | 0.896 | 0.852 |
| | 0.897 | 0.930 | 0.955 | 0.928 | 0.933 | 0.892 | 0.868 | 0.958 |
| | 0.904 | 0.920 | 0.943 | 0.918 | 0.924 | 0.929 | 0.902 | 0.962 |
| | 0.920 | 0.923 | 0.932 | 0.921 | 0.891 | 0.887 | 0.914 | 0.955 |
| | 0.913 | 0.917 | 0.922 | 0.923 | 0.870 | 0.892 | 0.938 | 0.935 |
| | 0.882 | 0.916 | 0.920 | 0.903 | 0.906 | 0.910 | 0.886 | 0.938 |
| | 0.904 | 0.922 | 0.948 | 0.918 | 0.918 | 0.893 | 0.885 | 0.951 |
| | 0.923 | 0.911 | 0.954 | 0.904 | 0.899 | 0.889 | 0.900 | 0.963 |
| | 0.917 | 0.937 | 0.937 | 0.908 | 0.929 | 0.896 | 0.915 | 0.972 |
| Average: | 0.895 | 0.922 | 0.916 | 0.914 | 0.892 | 0.891 | 0.903 | 0.935 |
| Max: | 0.923 | 0.940 | 0.955 | 0.946 | 0.933 | 0.929 | 0.943 | 0.972 |
| Min: | 0.792 | 0.894 | 0.815 | 0.871 | 0.846 | 0.837 | 0.857 | 0.852 |

Table A.3: Complete Gesture Recognition Accuracy Results for CIFF Subset$_5$

| Subject | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| CIFF DS$_5$ | 0.914 | 0.923 | 0.889 | 0.881 | 0.909 | 0.876 | 0.914 | 0.858 |
| | 0.932 | 0.917 | 0.890 | 0.867 | 0.878 | 0.914 | 0.933 | 0.849 |
| | 0.926 | 0.951 | 0.946 | 0.969 | 0.937 | 0.865 | 0.916 | 0.971 |
| | 0.886 | 0.941 | 0.847 | 0.962 | 0.917 | 0.953 | 0.950 | 0.944 |
| | 0.963 | 0.954 | 0.897 | 0.965 | 0.922 | 0.957 | 0.953 | 0.976 |
| | 0.815 | 0.957 | 0.794 | 0.904 | 0.972 | 0.962 | 0.956 | 0.876 |
| | 0.853 | 0.914 | 0.825 | 0.895 | 0.974 | 0.919 | 0.949 | 0.856 |
| | 0.935 | 0.958 | 0.814 | 0.873 | 0.946 | 0.959 | 0.949 | 0.875 |
| | 0.949 | 0.971 | 0.863 | 0.910 | 0.951 | 0.933 | 0.957 | 0.875 |
| | 0.866 | 0.932 | 0.914 | 0.924 | 0.869 | 0.848 | 0.910 | 0.948 |
| | 0.928 | 0.953 | 0.962 | 0.945 | 0.869 | 0.847 | 0.926 | 0.961 |
| | 0.939 | 0.964 | 0.814 | 0.964 | 0.911 | 0.877 | 0.938 | 0.971 |
| | 0.831 | 0.950 | 0.727 | 0.956 | 0.907 | 0.934 | 0.916 | 0.976 |
| | 0.934 | 0.927 | 0.889 | 0.913 | 0.906 | 0.934 | 0.939 | 0.959 |
| | 0.917 | 0.946 | 0.928 | 0.941 | 0.909 | 0.941 | 0.950 | 0.971 |
| | 0.936 | 0.976 | 0.847 | 0.926 | 0.980 | 0.867 | 0.955 | 0.864 |
| | 0.893 | 0.959 | 0.915 | 0.900 | 0.970 | 0.946 | 0.937 | 0.869 |
| | 0.890 | 0.954 | 0.919 | 0.852 | 0.967 | 0.928 | 0.951 | 0.866 |
| | 0.884 | 0.970 | 0.968 | 0.943 | 0.941 | 0.922 | 0.903 | 0.957 |
| | 0.928 | 0.968 | 0.969 | 0.958 | 0.897 | 0.867 | 0.938 | 0.955 |
| | 0.962 | 0.928 | 0.569 | 0.942 | 0.948 | 0.936 | 0.898 | 0.964 |
| | 0.935 | 0.950 | 0.951 | 0.893 | 0.975 | 0.920 | 0.891 | 0.979 |
| | 0.892 | 0.949 | 0.812 | 0.944 | 0.967 | 0.938 | 0.898 | 0.957 |
| | 0.865 | 0.969 | 0.911 | 0.959 | 0.921 | 0.941 | 0.978 | 0.979 |
| | 0.849 | 0.957 | 0.934 | 0.949 | 0.907 | 0.893 | 0.962 | 0.961 |
| Average: | 0.905 | 0.950 | 0.872 | 0.925 | 0.930 | 0.915 | 0.935 | 0.929 |
| Max: | 0.963 | 0.976 | 0.969 | 0.969 | 0.980 | 0.962 | 0.978 | 0.979 |
| Min: | 0.815 | 0.914 | 0.569 | 0.852 | 0.869 | 0.847 | 0.891 | 0.849 |

# Appendix B

# Multi-Label Gesture Recognition Results

Table B.1: Complete Gesture Recognition Accuracy Results for CIFF Subset$_9$

| Subject | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| CIFF DS$_9$ | 0.812 | 0.881 | 0.881 | 0.872 | 0.828 | 0.841 | 0.874 | 0.860 |
| | 0.808 | 0.885 | 0.861 | 0.856 | 0.852 | 0.817 | 0.843 | 0.854 |
| | 0.831 | 0.908 | 0.861 | 0.862 | 0.848 | 0.843 | 0.850 | 0.879 |
| | 0.822 | 0.861 | 0.872 | 0.866 | 0.852 | 0.841 | 0.838 | 0.898 |
| | 0.826 | 0.859 | 0.861 | 0.887 | 0.832 | 0.809 | 0.848 | 0.867 |
| | 0.822 | 0.846 | 0.889 | 0.844 | 0.847 | 0.862 | 0.848 | 0.900 |
| | 0.820 | 0.886 | 0.729 | 0.840 | 0.876 | 0.847 | 0.852 | 0.862 |
| | 0.844 | 0.893 | 0.885 | 0.855 | 0.869 | 0.830 | 0.855 | 0.879 |
| | 0.840 | 0.876 | 0.859 | 0.833 | 0.869 | 0.862 | 0.861 | 0.858 |
| | 0.820 | 0.911 | 0.857 | 0.847 | 0.873 | 0.830 | 0.861 | 0.870 |
| | 0.826 | 0.902 | 0.864 | 0.868 | 0.884 | 0.831 | 0.863 | 0.873 |
| | 0.811 | 0.876 | 0.876 | 0.846 | 0.890 | 0.880 | 0.869 | 0.882 |
| | 0.817 | 0.890 | 0.875 | 0.873 | 0.846 | 0.858 | 0.851 | 0.900 |
| | 0.832 | 0.878 | 0.851 | 0.882 | 0.860 | 0.837 | 0.869 | 0.881 |
| | 0.792 | 0.883 | 0.852 | 0.877 | 0.857 | 0.875 | 0.858 | 0.903 |
| Average: | 0.822 | 0.882 | 0.858 | 0.861 | 0.859 | 0.844 | 0.856 | 0.878 |
| Max: | 0.844 | 0.911 | 0.889 | 0.887 | 0.890 | 0.880 | 0.874 | 0.903 |
| Min: | 0.792 | 0.846 | 0.729 | 0.833 | 0.828 | 0.809 | 0.838 | 0.854 |

Table B.2: Complete Gesture Recognition Accuracy Results for CIFF Subset$_7$

| Subject | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| CIFF DS$_7$ | 0.809 | 0.870 | 0.810 | 0.827 | 0.823 | 0.809 | 0.842 | 0.846 |
| | 0.845 | 0.875 | 0.853 | 0.830 | 0.822 | 0.825 | 0.851 | 0.828 |
| | 0.836 | 0.870 | 0.863 | 0.881 | 0.814 | 0.843 | 0.846 | 0.864 |
| | 0.798 | 0.835 | 0.742 | 0.852 | 0.809 | 0.876 | 0.826 | 0.822 |
| | 0.796 | 0.905 | 0.861 | 0.861 | 0.832 | 0.866 | 0.857 | 0.895 |
| | 0.844 | 0.881 | 0.882 | 0.851 | 0.827 | 0.819 | 0.849 | 0.874 |
| | 0.828 | 0.844 | 0.828 | 0.843 | 0.790 | 0.818 | 0.869 | 0.897 |
| | 0.827 | 0.852 | 0.841 | 0.865 | 0.840 | 0.885 | 0.862 | 0.881 |
| | 0.858 | 0.840 | 0.838 | 0.873 | 0.874 | 0.850 | 0.862 | 0.835 |
| | 0.755 | 0.882 | 0.875 | 0.888 | 0.835 | 0.801 | 0.829 | 0.871 |
| | 0.793 | 0.800 | 0.852 | 0.841 | 0.806 | 0.843 | 0.815 | 0.879 |
| | 0.799 | 0.880 | 0.804 | 0.843 | 0.814 | 0.823 | 0.830 | 0.835 |
| | 0.823 | 0.899 | 0.884 | 0.868 | 0.865 | 0.833 | 0.832 | 0.909 |
| | 0.818 | 0.872 | 0.813 | 0.878 | 0.875 | 0.857 | 0.843 | 0.913 |
| | 0.825 | 0.869 | 0.884 | 0.821 | 0.791 | 0.825 | 0.886 | 0.876 |
| | 0.810 | 0.833 | 0.868 | 0.850 | 0.803 | 0.853 | 0.850 | 0.865 |
| | 0.833 | 0.830 | 0.771 | 0.853 | 0.823 | 0.836 | 0.833 | 0.906 |
| | 0.809 | 0.893 | 0.856 | 0.824 | 0.861 | 0.815 | 0.838 | 0.895 |
| | 0.847 | 0.862 | 0.859 | 0.849 | 0.825 | 0.841 | 0.868 | 0.817 |
| | 0.850 | 0.893 | 0.878 | 0.836 | 0.876 | 0.831 | 0.878 | 0.926 |
| Average: | 0.820 | 0.864 | 0.843 | 0.852 | 0.830 | 0.837 | 0.848 | 0.872 |
| Max: | 0.858 | 0.905 | 0.884 | 0.888 | 0.876 | 0.885 | 0.886 | 0.926 |
| Min: | 0.755 | 0.800 | 0.742 | 0.821 | 0.790 | 0.801 | 0.815 | 0.817 |

Table B.3: Complete Gesture Recognition Accuracy Results for CIFF $\text{Subset}_5$

| Subject | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| CIFF $\text{DS}_5$ | 0.786 | 0.854 | 0.771 | 0.784 | 0.806 | 0.793 | 0.815 | 0.840 |
| | 0.788 | 0.871 | 0.815 | 0.733 | 0.783 | 0.850 | 0.837 | 0.836 |
| | 0.803 | 0.883 | 0.788 | 0.838 | 0.818 | 0.790 | 0.860 | 0.874 |
| | 0.756 | 0.850 | 0.600 | 0.869 | 0.849 | 0.819 | 0.779 | 0.881 |
| | 0.791 | 0.798 | 0.761 | 0.867 | 0.802 | 0.850 | 0.803 | 0.892 |
| | 0.792 | 0.887 | 0.764 | 0.842 | 0.827 | 0.777 | 0.788 | 0.748 |
| | 0.720 | 0.848 | 0.630 | 0.793 | 0.864 | 0.719 | 0.838 | 0.809 |
| | 0.838 | 0.815 | 0.735 | 0.732 | 0.860 | 0.816 | 0.858 | 0.786 |
| | 0.819 | 0.763 | 0.755 | 0.808 | 0.840 | 0.804 | 0.827 | 0.798 |
| | 0.734 | 0.879 | 0.785 | 0.802 | 0.724 | 0.792 | 0.811 | 0.817 |
| | 0.795 | 0.848 | 0.848 | 0.880 | 0.802 | 0.781 | 0.827 | 0.859 |
| | 0.767 | 0.864 | 0.815 | 0.920 | 0.773 | 0.772 | 0.829 | 0.906 |
| | 0.771 | 0.854 | 0.694 | 0.852 | 0.758 | 0.875 | 0.814 | 0.910 |
| | 0.815 | 0.876 | 0.776 | 0.854 | 0.824 | 0.809 | 0.802 | 0.865 |
| | 0.711 | 0.743 | 0.700 | 0.803 | 0.839 | 0.712 | 0.799 | 0.845 |
| | 0.787 | 0.856 | 0.772 | 0.838 | 0.884 | 0.801 | 0.811 | 0.829 |
| | 0.758 | 0.678 | 0.812 | 0.787 | 0.863 | 0.857 | 0.853 | 0.737 |
| | 0.782 | 0.870 | 0.787 | 0.776 | 0.845 | 0.847 | 0.764 | 0.837 |
| | 0.714 | 0.830 | 0.729 | 0.833 | 0.876 | 0.779 | 0.785 | 0.881 |
| | 0.845 | 0.848 | 0.833 | 0.814 | 0.766 | 0.791 | 0.825 | 0.898 |
| | 0.800 | 0.863 | 0.799 | 0.847 | 0.807 | 0.854 | 0.812 | 0.887 |
| | 0.778 | 0.858 | 0.808 | 0.866 | 0.898 | 0.767 | 0.794 | 0.939 |
| | 0.775 | 0.885 | 0.723 | 0.855 | 0.906 | 0.824 | 0.802 | 0.874 |
| | 0.705 | 0.874 | 0.796 | 0.886 | 0.833 | 0.813 | 0.797 | 0.867 |
| | 0.710 | 0.835 | 0.764 | 0.819 | 0.746 | 0.726 | 0.852 | 0.859 |
| Average: | 0.774 | 0.841 | 0.762 | 0.828 | 0.824 | 0.801 | 0.815 | 0.851 |
| Max: | 0.845 | 0.887 | 0.848 | 0.920 | 0.906 | 0.875 | 0.860 | 0.939 |
| Min: | 0.705 | 0.678 | 0.600 | 0.732 | 0.724 | 0.712 | 0.764 | 0.737 |