

ROBUST AND RELIABLE HAND GESTURE
RECOGNITION FOR MYOELECTRIC CONTROL

A THESIS SUBMITTED TO
THE UNIVERSITY OF KENT
IN THE SUBJECT OF COMPUTER SCIENCE
FOR THE DEGREE
OF DOCTOR OF PHILOSOPHY.

By
Yuzhou Lin
August 2023

Abstract

Surface Electromyography (sEMG) is a physiological signal to record the electrical activity of muscles by electrodes applied to the skin. In the context of Muscle-Computer Interaction (MCI), systems are controlled by transforming myoelectric signals into interaction commands that convey the intent of user movement, mostly for rehabilitation purposes. Taking the myoelectric hand prosthetic control as an example, using sEMG recorded from the remaining muscles of the stump can be considered as the most natural way for amputees who lose their limbs to perform activities of daily living with the aid of prostheses. Although the earliest myoelectric control research can date back to the 1950s, there still exist considerable challenges to address the significant gap between academic research and industrial applications. Most recently, pattern recognition-based control is being developed rapidly to improve the dexterity of myoelectric prosthetic devices due to the recent development of machine learning and deep learning techniques.

It is clear that the performance of Hand Gesture Recognition (HGR) plays an essential role in pattern recognition-based control systems. However, in reality, the tremendous success in achieving very high sEMG-based HGR accuracy ($\geq 90\%$) reported in scientific articles produced only limited clinical or commercial impact. As many have reported, its real-time performance tends to degrade significantly as a result of many confounding factors, such as electrode shift, sweating, fatigue, and day-to-day variation. The main interest of the present thesis is, therefore, to improve the robustness of sEMG-based HGR by taking advantage of the most

recent advanced deep learning techniques to address several practical concerns. Furthermore, the challenge of this research problem has been reinforced by only considering using raw sparse multichannel sEMG signals as input.

Firstly, a framework for designing an uncertainty-aware sEMG-based hand gesture classifier is proposed. Applying it allows us to quickly build a model with the ability to make its inference along with explainable quantified multidimensional uncertainties. This addresses the black-box concern of the HGR process directly.

Secondly, to fill the gap of lacking consensus on the definition of model reliability in this field, a proper definition of model reliability is proposed. Based on it, reliability analysis can be performed as a new dimension of evaluation to help select the best model without relying only on classification accuracy. Our extensive experimental results have shown the efficiency of the proposed reliability analysis, which encourages researchers to use it as a supplementary tool for model evaluation.

Next, an uncertainty-aware model is designed based on the proposed framework to address the low robustness of hand grasp recognition. This offers an opportunity to investigate whether reliable models can achieve robust performance. The results show that the proposed model can improve the long-term robustness of hand grasp recognition by rejecting highly uncertain predictions.

Finally, a simple but effective normalisation approach is proposed to improve the robustness of inter-subject HGR, thus addressing the clinical challenge of having only a limited amount of data from any individual. The comparison results show that better performance can be obtained by it compared to a state-of-the-art (SoA) transfer learning method when only one training cycle is available.

In summary, this study presents promising methods to pursue an accurate, robust, and reliable classifier, which is the overarching goal for sEMG-based HGR. The direction for future work would be the inclusion of these in real-time myoelectric control applications.

Copyright Declaration

The copyright for this thesis lies with the author and is available under the Creative Commons Attribution Noncommercial No Derivatives Licence. Researchers are free to copy or share the thesis when they attribute it, do not use it for commercial purposes, and do not modify or build upon it for whatever purpose. Researchers must make clear to other users the licence terms of this work before reuse or redistribution.

Declaration of Originality

I, Yuzhou Lin, hereby declare that the work presented in this PhD thesis entitled ‘Robust and reliable hand gesture recognition for myoelectric control’ is original and has not been submitted or published in any other form. The research and analysis presented in this thesis are the result of my own efforts and have not been taken from any other source, except where explicitly acknowledged in the text. Any ideas or theories that have been used from the work of others have been properly referenced and cited in accordance with academic standards. I take full responsibility for any errors or omissions that may have occurred in the preparation of this thesis. This declaration is made with the understanding that any breach of originality will result in serious consequences, including but not limited to, the invalidation of the degree awarded and disciplinary action.

Acknowledgements

First and foremost, I would like to thank my main supervisor, Professor Caroline Li, for her guidance, insight, enthusiasm and patience throughout my graduate career. I appreciate the freedom I was given to design and develop my own project, especially under the conditions of the Covid-19 pandemic. Furthermore, I would like to thank my second supervisor, Dr Palaniappan Ramaswamy, for his technical and moral support through all the stages of this degree. I am grateful that he has always made time to address my questions and has provided honest, candid, and wise advice. Finally, I would like to express my sincere appreciation and thankfulness to my third supervisor, Professor Philippe De Wilde, for his valuable comments and discussions while writing various papers and this thesis. Thanks to his deep and systematic knowledge and abundant research experience, I can always learn something from our conversation.

The school of computing Medway research group was such a caring and friendly team to be part of. To our lovely and enthusiastic administration staff, Janine Jarvis and Shannon Croft; to my dear colleagues, James Brookhouse, Ayah Mohamed Helal, Ismail M. Anwar, George Langroudi, Constantinos Patlatzoglou, Riku Ihalainen, Lam Pham, together with other visiting researchers, Donggeon Lee, Xiaoxiao Miao and Helen Costa Lima: you were my family on this side of the world. Thank Prof. Ian McLoughlin and Dr Kwai Yoke McLoughlin for opening your home to us to celebrate the Chinese new year together.

I would also like to thank my friends, who have been a constant source of laughter and joy during the most challenging times. Your friendship has brought colour and light into my life, and I am so grateful for all the memories we have shared together. I would also like to extend my sincere gratitude to all of the researchers and colleagues who provided me with the opportunity to engage in thought-provoking discussions. These interactions allowed me to gain new insights and perspectives that have greatly enriched my understanding of the subject matter.

Finally, I am especially grateful to my parents, who have always been there for me, both emotionally and financially, through thick and thin. Their love, sacrifice, and guidance have been instrumental in shaping me into the person I am today, and I am forever grateful for their unwavering support and belief in me.

Contents

Bibliography	i
Abstract	ii
Copyright Declaration	iv
Declaration of Originality	v
Acknowledgements	vi
Contents	viii
List of Tables	xiii
List of Figures	xv
List of Algorithms	xix
Abbreviations	1
1 Introduction	6
1.1 Background and Motivation	6
1.2 sEMG-based HGR Overview	7
1.2.1 Data Acquisition	8
1.2.2 Data Preprocessing	9

1.2.3	Feature Extraction	10
1.2.4	Pattern Recognition	11
1.2.5	Post-processing	12
1.2.6	Performance Evaluation	13
1.3	Problem Statement	14
1.3.1	Practical considerations	14
1.3.2	Technical challenges	16
1.4	Aims and Objectives	16
1.5	Contributions	17
1.6	Publications	19
1.7	Thesis Structure	21
2	Literature Review	22
2.1	Electromyography	22
2.1.1	Discovery History	22
2.1.2	Biological Basis	24
2.1.3	Influence factors	28
2.1.4	Characteristics of sEMG	31
2.1.5	Applications	31
2.2	sEMG-based Hand Gesture Recognition	35
2.2.1	Evolution from static to dynamic HGR	35
2.2.2	Advancement from feature engineering to feature learning	37
2.2.3	Transition from accuracy to robustness	39
2.3	Summary	41
3	Methods	42
3.1	Introduction	42
3.2	Subjective Logic	43
3.3	Evidential Deep Learning	45

3.4	Evidential Convolutional Neural Network	46
3.5	Loss functions	48
3.6	Hyperparameters	51
3.7	Evidential uncertainty	51
3.8	Summary	52
4	Reliability Analysis of Finger Movement Recognition	53
4.1	Introduction	53
4.2	Problem Statement	56
4.3	Database	57
4.4	Models	58
4.5	Illustration	61
4.6	Experimental Setup	66
4.7	Performance Evaluation	70
	4.7.1 Evaluation of Accuracy	70
	4.7.2 Evaluation of Reliability	71
	4.7.3 Evaluation under Cross-Validation	72
4.8	Results	72
	4.8.1 Accuracy Analysis	72
	4.8.2 Reliability Analysis	77
4.9	Discussion	79
4.10	Summary	82
5	Improve the Long-Term Robustness of Hand Grasp Recognition	83
5.1	Introduction	83
5.2	Related Work	85
5.3	Multidimensional uncertainty-aware models	88
	5.3.1 Baseline model	88
	5.3.2 Uncertainty-aware models	89

5.4	Experiments	91
5.4.1	Data Preprocessing	91
5.4.2	Model training	92
5.4.3	Rejection-capable performance evaluation	95
5.4.4	Recognition Performance - No rejections	96
5.4.5	Recognition Performance - With Rejections	98
5.4.6	Comparison with SoA	104
5.5	Discussion	104
5.6	Conclusion	110
6	Improve the Inter-subject recognition with normalisation	111
6.1	Introduction and related work	111
6.2	Datasets	113
6.2.1	<i>G. Dataset</i>	113
6.2.2	<i>M. Dataset</i>	114
6.3	Hypotheses	114
6.4	Methods	117
6.4.1	Referencing normalisation	117
6.4.2	The model architecture	118
6.4.3	System structure	119
6.5	Results	120
6.6	Conclusion	122
7	Conclusions and Future Research	124
7.1	Conclusions	124
7.2	Directions for Future research	129
7.2.1	Multimodal fusion for improving model reliability	130
7.2.2	Uncertainty-aware model design to extract advanced spatial-temporal features	131

7.2.3	Improve the explicit training of ECNN	134
7.2.4	Improve the predictive uncertainties	134
7.2.5	Improve inter-subject recognition with uncertainty	135
7.2.6	Real-time myoelectric control interfaces	136
Bibliography		137
A Hand movements demonstrations		162
B The Determined Optimal Hyperparameters of Models employed for reliability analysis of finger movement recognition		167

List of Tables

1	Hyperparameter Search Space	69
2	Macroaverage Recall of the ConvNets with Comparisons	73
3	Reliability Comparison of the ConvNets by Evaluating the Misclassification Detection regarding Uncertainty Estimates	78
4	Summary of the research work on NinaPro DB6 in the literature.	87
5	The Determined Optimal Hyperparameters of 2D CNN and 2D ECNNs on <i>NinaPro DB6</i>	93
6	The Determined Optimal Hyperparameters of 3D CNN and 3D ECNNs on <i>NinaPro DB6</i>	94
7	Inter-Session Cross-Day Standard Recognition Accuracy of ConvNets with Comparisons	97
8	Overall Rejection-capable Performance Comparison of 2D ConvNets	99
9	Overall Rejection-Capable Performance Comparison of 3D ConvNets	100
10	Comparison of the proposed models with the state-of-the-art (SoA) results on the inter-session cross-day accuracy	103
11	Reliability Comparison of 2D and 3D ConvNets by Evaluating the Misclassification Detection regarding Uncertainty Estimates	105
12	Description of muscle location and sEMG electrode placement in <i>G. Dataset</i>	113
13	Proposed CNN architecture for <i>G. Dataset</i>	118

14	Classification accuracy of the CNNs on three datasets with respect to the normalisation methods applied	120
15	Classification accuracy of CNNs on three datasets with raw data and transfer learning	121
16	The Determined Optimal Hyperparameters of CNN with Stratified Nested Cross-Validation on <i>NinaPro DB5</i>	168
17	The Determined Optimal Hyperparameters of ECNN-A with Stratified Nested Cross-Validation on <i>NinaPro DB5</i>	169
18	The Determined Optimal Hyperparameters of ECNN-B with Stratified Nested Cross-Validation on <i>NinaPro DB5</i>	170
19	The Determined Optimal Hyperparameters of ECNN-C with Stratified Nested Cross-Validation on <i>NinaPro DB5</i>	171

List of Figures

1	The typical pipeline of Surface Electromyography (sEMG)-based Hand Gesture Recognition (HGR).	8
2	The high level structure of this thesis.	20
3	A brief history of the discovery of Electromyography.	23
4	The outline of the decomposition of the sEMG signal into its constituent Motor Unit Action Potentials (MUAPs). (Luca et al. 2006)	25
5	The illustration of how the action potential of each fiber contributes to the Electromyography (EMG) signal. (Kamen 2013) There are 5 muscle fibers in this example as labelled in the left panel while their corresponding action potentials are shown in the right one. Furthermore, two Motor Units (MUs), α_A and α_B , are shown here. The amplitude of each MUAP is represented as the algebraic sum of action potentials of the contributed muscle fibers. Eventually, the recorded EMG signal is the algebraic sum of all MUs.	26
6	The illustration of mathematical model of EMG signal. (Akhmadeev 2019)	27
7	The illustration of Myo armband sensor by Thalmic labs, which has ceased production recently and no corresponding website is available now. The bottom panel has shown all five possible hand gestures that can be classified from built-in programs (Tatarian et al. 2018).	33

8	The detailed illustrations of the proposed Evidential Convolutional Neural Network (ECNN) and its conventional version (i.e., Convolutional Neural Network (CNN)). BN refers to Batch Normalisation. PReLU refers to Parametric Rectified Linear Unit. Conv_X (A×B) refers to the Xth convolution layer with (A×B) kernel valid padding. FC_X refers to the Xth Fully-Connected layer.	60
9	Sequential predictions of the ‘thumb adduction’ (class 10) on offline testing. Note that the sequential predictions of a wrong class are presented only if the predictive probability for at least one of them is above 0.5. Note that the ECNN refers to ECNN-A in this figure. Top: The predicted probabilities of the CNN; Class 3 and 12 refer to ‘middle flexion’ and ‘thumb flexion’. Middle: The predicted probabilities of the ECNN with its evidential uncertainty. Bottom: The sum of moving averages of 16-channel raw rectified sEMG signals with absolute values regarding the dynamic finger movement of ‘thumb adduction’. The u.a. denotes ‘unitless’ activation since sEMG recorded by Thalmic Myo armbands is claimed to be ‘unitless’ with an unknown conversion from mV.	62
10	Three comparison results of classifying 12 finger movements between CNN and ECNN when rejecting uncertain predictions. Note that the ECNN refers to ECNN-A in this figure.	65
11	Accuracy comparisons of the CNN and ECNN variants with nested cross validation. 0 is CNN and 1 – 3 refer to ECNN-A, ECNN-B and ECNN-C.	74

12	Confusion matrices averaged over all subjects with nested cross validation of the CNN and ECNN variants. Note that 0, 2, 4, 6, 10 are the flexion movements of finger Index, Middle, Ring, Little finger and Thumb; The extension movements of each finger are then represented as class 1, 3, 5, 7, 11. Furthermore, 8, 9 refer to Thumb adduction and abduction.	75
13	The mean ARC plots of all models under cross-validation (CV) scheme when considering the ‘overall’ uncertainty estimate. . . .	76
14	The model structure of Evidential Convolutional Networks (ECNN) and its conventional version (i.e., CNN).	90
15	The confusion matrix of misclassification detection. True Acceptance Rate (TAR) can be considered the accuracy of rejection-capable sEMG-based hand grasp recognition and Precision is actually the True Rejection Rate (TRR) defined in this chapter. . . .	96
16	Multidimensional uncertainty-based rejection-capable accuracy in several data acquisitions across days	102
17	The normalised confusion matrices of 2D ECNN-A and 2D ECNN-B averaged for 10 runs over 8 subjects. Note that an extra x-axis was introduced when presenting confusion matrix with rejection, which should be considered a separated part from the standard one where each cell refers to the rejection rate for a class.	108
18	The normalised confusion matrices of 3D ECNN-A and 3D ECNN-B averaged for 10 runs over eight subjects. Note that an extra x-axis was introduced when presenting confusion matrix with rejection, which should be considered a separated part from the standard one where each cell refers to the rejection rate for a class.	109
19	Probability density functions of the sEMG of the same session but different subjects	115

20	Probability density functions of the sEMG from the same subject but different sessions	116
21	A high-level diagram illustrating the steps for inter-subject sEMG-based hand gesture classification	119
22	An illustration of different multimodal fusion strategies. (a) Early or data-level fusion, (b) late or decision-level fusion, and (c) intermediate fusion. (Ramachandram and Taylor 2017)	131
23	A collage comprised of all available pictures from (Francesca Palermo 2017) that contain the information of electrode placement for NinaPro DB6	132
24	A graph network representing the physical spatial dependency of 14 electrodes used in NinaPro DB6	133
25	The 12 basic finger movements considered in the NinaPro Database 5, Exercise A, with corrected labels (Atzori and Müller 2015). . .	163
26	The 7 hand grasps with their relative holding objects considered in the NinaPro Database 6 (Palermo et al. 2017).	164
27	A subset of hand gesture examples from G. Dataset (Li et al. 2011).	165
28	The 7 commonly seen hand gestures considered in the M. Dataset (Côté-Allard et al. 2019).	166

List of Algorithms

1	Model Training with Stratified Nested CV	67
---	--	----

Abbreviations

ADAM Adaptive Moment Estimation 61, 92, 118

ADL Activities of Daily Living 36, 83

ANN Artificial Neural Network 38

AP Average Precision 71

AR Auto-regressive 11

ARC Accuracy-Rejection Curve 70, 76, 82, 125

AUPRC Area Under Precision-Recall Curve 56, 57, 71, 72, 81

AUROC Area Under Receiver Operating Characteristic 56, 57, 71, 81

AWGN Additive White Gaussian Noise 26

BF Bayesian Fusion 13

BN Batch Normalisation 59, 118, 121

CCA Canonical Correlation Analysis 112

CNN Convolutional Neural Network xvi, xvii, 12, 37, 40, 42, 47, 54–56, 58–66, 69, 70, 72–77, 79, 80, 82, 84, 85, 87–89, 96–98, 103–105, 112, 118–122, 125, 126, 129–132, 135

CV cross-validation xvii, 56, 66, 68–70, 72, 73, 76, 81

CWT Continuous Wavelet Transform 11

DST Dempster-Shafer Theory of Evidence 43

ECNN Evidential Convolutional Neural Network xvi, xvii, 18, 42, 43, 47–52, 55, 56, 59–66, 69, 70, 72–77, 79–82, 85, 88, 89, 96–98, 101, 103–107, 110, 125–129, 131, 134, 135

EDL Evidential Deep Learning 42, 45–47, 52, 59, 85, 89, 104, 125, 133

EEG Electroencephalogram 6, 34, 130

EMG Electromyography xv, 6, 7, 22–32, 116

FAR False Acceptance Rate 64

FC Frequency Ratio 11

FD Frequency Domain 10, 11

FFT Fast Fourier Transform 87

FN False Negative 95

FRR False Rejection Rate 64

GCN Graph Convolutional Network 89, 132

GMM Gaussian Mixture Model 11

GRU Gated Recurrent Unit 12, 37

HCI Human-Computer Interaction 6, 16, 31, 32, 51

HD high-density 8

HGR Hand Gesture Recognition ii, iii, xv, 6–8, 10–18, 21, 22, 35–42, 47, 48, 50, 51, 53–55, 61, 63, 64, 81–85, 88, 106, 111, 114, 117, 119, 120, 122–130, 133–135

HMI Human-Machine Interaction 83, 84

HMM Hidden Markov Model 11

HPO Hyperparameter Optimisation 51, 52, 68, 69, 92

IMU Inertial Measurement Unit 130

KL Kullback-Leibler 49, 50, 69, 80

kNN k -nearest neighbor 12, 38

LDA Linear Discriminant Analysis 12, 36

LOSOCV Leave-One-Subject-Out Cross-Validation 112, 120, 122, 126, 129

LSTM Long Short-Term Memory 12, 37, 38

MAV Mean Absolute Value 10, 87

MCI Muscle-Computer Interaction ii, 33

MDF Median Frequency 11

mDWT marginal Discrete Wavelet Transform 87

MNF Mean Frequency 11

MU Motor Unit xv, 24–26

MUAP Motor Unit Action Potential xv, 24–26, 28

MV Majority Voting 13

- MV-CNN** Multi-View Convolutional Neural Network 87
- nAUPRC** normalised Area Under Precision-Recall Curve 56, 71, 81, 105, 125
- NB** Naïve Bayes 12
- NinaPro** Non-Invasive Adaptive Hand Prosthetics 38, 39, 57, 58, 61, 68, 73, 80, 85, 86, 89, 91, 104, 110, 126, 128, 132
- PDF** Probability Density Function 43–45, 115
- PRC** Precision-Recall Curve 71, 81
- PReLU** Parametric Rectified Linear Unit 59, 118
- PS** Power Spectrum 11
- RF** Random Forest 12, 38
- RNN** Recurrent Neural Network 37
- ROC** Receiver Operating Characteristic 81
- RR** Rejection Rate 64, 66, 70, 76, 80, 81, 98, 106, 107
- sEMG** Surface Electromyography ii, iii, xv, xvi, 7–18, 21, 22, 25, 28–42, 47, 50–59, 61–64, 66–68, 70, 71, 73, 80–85, 87–89, 91, 92, 96, 98, 101, 103, 104, 106, 110–117, 119, 120, 122–135
- SL** Subjective Logic 42–45, 47, 89
- SLR** Sign Language Recognition 33
- SNR** Signal-to-Noise Ratio 29, 30, 86
- SoA** state-of-the-art iii, xiii, 17, 22, 35, 38, 41, 51, 54, 56, 68, 80, 82, 84, 88, 92, 103, 104, 110, 112, 122, 126, 128, 129

SSC Slope Sign Change 10

STFT Short Time Fourier Transform 31

SVM Support Vector Machine 12, 38, 112

TAR True Acceptance Rate xvii, 64, 70, 81, 95, 96, 98, 101, 105–107

TCN Temporal Convolutional Network 12, 37, 38, 88, 103, 133

TD Time Domain 10, 11

TFD Time-Frequency Domain 10, 11

TL Transfer Learning 112, 121–123, 126, 129

TN True Negative 95

TP True Positive 95

TPE Tree-structured Parzen Estimator 56, 68, 92

TRR True Rejection Rate xvii, 64, 66, 81, 96, 98, 105–107

VAR Variance 87

WL Waveform Length 10

WT Wavelet Transform 31

ZC Zero Crossing 10

Chapter 1

Introduction

1.1 Background and Motivation

Gestures, particularly hand gestures, are the second most widely used type of interaction in everyday life after speech. It plays a crucial role in Human-Computer Interaction (HCI) systems. A hand gesture is commonly defined as the physical movement of the fingers, hand, and arm. It is clear that Hand Gesture Recognition (HGR) remains an important link to HCI because effective identification of hand gestures can directly contribute to the flow of meaningful information from users to devices. This can help HCI establish an intuitive and interactive human-computer learning environment from which user input is provided easily and naturally.

Generally, the observable manifestations of gestures used for recognition can be captured by vision- and non-vision-based sensors. Vision-based HGR requires cameras mainly, whereas non-vision-based one decodes the intention of hand motion by sensing physiological signals such as Electroencephalogram (EEG) and Electromyography (EMG). Concisely, EEG and EMG capture electrical activity from the brain and muscles, respectively. Leaving aside any discussion of sensor selection, which depends on the scope of the application, one cannot deny that

HGR based on physiological signals, especially EMG, is a natural or perhaps only choice for rehabilitation applications required for intuitive control by people with certain disabilities. For example, it is natural to control a prosthetic hand based on EMG collected from the remaining muscles for patients who have their hands amputated. Furthermore, it is clear that connecting dynamic hand gestures with control commands can lead to various and broad applications in more general areas such as robot manipulation, computer games, virtual reality, and smart home.

Recently, HGR based on Surface Electromyography (sEMG) has been extensively investigated by many researchers, where sEMG refers to the collective electrical signals of muscles collected by non-invasive electrodes. It has been used to control the hand (Finley and Wirta 1967) since the late 1960s and has been increasingly used to control robotic prostheses and exoskeletons over the last decade. This is mainly due to the non-invasive characterisation of sEMG signals, which has shown great commercial advantages over traditional invasive methods. However, the potential of myoelectric control, i.e., controlled by sEMG, can be severely limited by the low robustness reported in the recent literature. Therefore, the development of robust and reliable HGR for myoelectric control is a highly relevant research priority.

1.2 sEMG-based HGR Overview

As a recognition problem, the typical pipeline of sEMG-based HGR includes six phases in the proper order, shown in Figure 1. Here, only the relevant fundamental information for each phase is mentioned for the purpose of providing a general background of sEMG-based HGR.

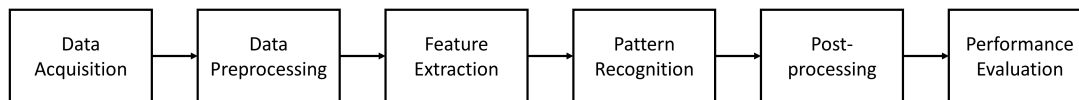


Figure 1: The typical pipeline of sEMG-based HGR.

1.2.1 Data Acquisition

There are two categories of sEMG signals, sparse multichannel sEMG (Atzori et al. 2014; Atzori and Müller 2015; Pizzolato et al. 2017) and high-density (HD)-sEMG (Rojas-Martínez, Mañanas and Alonso 2012; Geng et al. 2016), from the perspective of the electrode density used during data acquisition. Sparse multichannel sEMG signals are collected by a limited number of electrodes attached to muscles based on an anatomic location strategy. Relatively, HD-sEMG signals are recorded by a large number of electrodes that are generally spatially and densely distributed around the forearm. Unfortunately, to our knowledge, there is no clear clarification on how many electrodes can be considered HD-sEMG. According to the sEMG datasets published so far, the number of electrodes used in the HD-sEMG and sparse multichannel datasets are usually ≥ 128 and ≤ 16 , respectively, and HD-sEMG signals are often recorded using two-dimensional electrode arrays.

Although HGR based on HD-sEMG can directly improve recognition accuracy due to the large amount of data collected, it requires additional computing efforts and processing facilities. In other words, it can impose higher hardware requirements, leading to increased power consumption and high costs. If the recognition performance of HGR with sparse multichannel sEMG signals can be improved, it will result in a more efficient approach that can be implemented with low-cost hardware resources. Therefore, this thesis aims to enhance HGR, particularly for rehabilitation applications, by only focussing on sparse multichannel sEMG signals despite the added difficulty. Note that, unless otherwise stated, sEMG-based HGR refers only to HGR with the sparse multichannel sEMG in this thesis.

1.2.2 Data Preprocessing

Generally, the collected sEMG signals need to be pre-processed to remove noise or reduce non-stationarity. Moreover, pre-processing techniques can help extract serviceable information from sEMG for effective analysis or feature extraction. Practical preprocessing techniques include, but are not limited to, relabelling, smoothing, normalisation, and segmentation (Li, Shi and Yu 2021).

Relabelling is used to synchronise the movements performed by a participant with the stimuli proposed by the acquisition software (Atzori, Cognolato and Müller 2016), which can be implemented offline with a generalised likelihood ratio algorithm (Kuzborskij, Gijsberts and Caputo 2012).

Smoothing is a widely used technique to counter the non-stationarity of sEMG signals by means of rectification, which involves calculating the absolute values of the signals and discarding any negative magnitudes. (Geng et al. 2016). To further smooth the electrical activities of the muscles, a low-pass Butterworth filter may be a good option in filter design (Atzori, Cognolato and Müller 2016; Geng et al. 2016; Rahimian et al. 2021). Furthermore, power line interference can be eliminated by a notch filter at 50Hz or 60Hz, depending on the location of data acquisition.

Normalisation is one of the possible approaches to reduce the variability of sEMG. It is known that sEMG is highly variable and can be affected by many factors, such as electrode placement (Jensen, Vasseljen and Westgaard 1993; Hargrove, Englehart and Hudgins 2006), skin-electrode contact impedance (Kaczmarek, Mańkowski and Tomczyński 2019), limb position (Scheme et al. 2010), perspiration (Abdoli-Eramaki et al. 2012), skin temperature (Winkel and Jørgensen 1991), and muscle fatigue (Wan et al. 2010). Lehman and McGill (1999) have reinforced the need for normalisation and recommended two normalisation approaches, that is, expressing sEMG as a percentage of an isometric maximum

voluntary contraction or a reference voluntary contraction. Furthermore, normalisation approaches can be applied to reduce the variations between subjects in the sEMG signals to improve subject-independent HGR accuracy (Wahid et al. 2018; Kerber, Puhl and Krüger 2017). Certainly, the choice of the normalisation method determines the interpretation of the sEMG signals by changing their amplitude and pattern (Burden 2010; Halaki and Gi 2012). Unfortunately, there is no universal rule to normalise sEMG until now.

To consider a real-time myoelectric control system to control rehabilitation devices or assistive robots, the sEMG signals must first be segmented with a sliding window for offline training and testing. Using a larger window could lead to longer processing delays and introduce more non-stationarity, whereas a shorter window may result in degraded recognition performance due to the limited information available. Previous research has suggested that the length of the sliding window should not exceed 300 ms (Hudgins, Parker and Scott 1993). More specifically, Englehart and Hudgins (2003) have shown that 150 – 250 ms windows are the optimal choices for sEMG sensors. Unlike disjointing windows, overlapping ones are preferred because they can generate fast and dense decision flows.

1.2.3 Feature Extraction

The main goal of feature extraction is to increase the information density implicit in sEMG signals, thus expecting to increase the distinction between gestures. The discriminative characteristics of the sEMG signals can be extracted by designing handcrafted features that fall into one of three categories in general: Time Domain (TD), Frequency Domain (FD), and Time-Frequency Domain (TFD) (Oskoei and Hu 2007). The TD features have the advantage of low computational complexity. One of the most commonly used traditional TD feature sets consists of Mean Absolute Value (MAV), Waveform Length (WL), Slope Sign Change (SSC), and Zero Crossing (ZC) (Hudgins, Parker and Scott 1993). The FD features refer to

features that describe spectral properties, which may be more robust than the TD ones (Gu et al. 2018). Common features of FD include Frequency Ratio (FC) (Os-koei and Hu 2007), Mean Frequency (MNF) (Thongpanja et al. 2013), Median Frequency (MDF) (Thongpanja et al. 2013), and Power Spectrum (PS) (Bao et al. 2022). TFD features deemed a significant approach to feature extraction as they examine the energy of the sEMG signals in both the time and frequency domains (Karlsson, Yu and Akay 2000; Côté-Allard et al. 2019). They often employ the wavelet transform, such as the Continuous Wavelet Transform (CWT), in their calculations. Furthermore, the characteristics of sEMG can be extracted from the perspective of time series modelling. For example, coefficients of the fourth order Auto-regressive (AR) were suggested to be used as a feature vector based on previous research work (Paiss and Inbar 1987; Phinyomark, Phukpattaranont and Limsakul 2012).

In contrast to the traditional hand-crafted features presented above, the features of sEMG can be learnt using deep learning, which is a multilevel representation learning approach. Feature learning has the advantages of being less labour-intensive and time-consuming since it does not require prior knowledge for feature determination and extra computation for feature extraction. Unlike taking manually extracted features as input to deep learning models, feature learning requires taking raw sEMG as input directly, which is called end-to-end learning.

1.2.4 Pattern Recognition

Extracted or learnt features are required to be classified into one of the predefined hand gestures. An ideal sEMG-based hand gesture classifier is expected to be accurate, robust, and reliable. Traditional HGR approaches rely on fuzzy theory (Chan et al. 2000; Karlik, Tokhi and Alci 2003; Ajiboye and Weir 2005), probabilistic models such as Gaussian Mixture Model (GMM) (Fukuda et al. 2003; Huang et al. 2005) and Hidden Markov Model (HMM) (Chan and Englehart 2005)

or machine learning algorithms such as Linear Discriminant Analysis (LDA) (Hargrove et al. 2010; Scheme and Englehart 2011; Scheme, Hudgins and Englehart 2013; Phinyomark et al. 2013; Wahid et al. 2018; Campbell, Phinyomark and Scheme 2019; Li et al. 2019; Kaczmarek, Mańkowski and Tomczyński 2019), Random Forest (RF) (Atzori et al. 2014; Al-Timemy et al. 2016; Wahid et al. 2018), Naïve Bayes (NB) (Chen et al. 2007; Scheme and Englehart 2011; Lu et al. 2017; Wahid et al. 2018), k -nearest neighbor (kNN) (Scheme and Englehart 2011; Atzori et al. 2014; Wahid et al. 2018; Li et al. 2019; Kaczmarek, Mańkowski and Tomczyński 2019) and Support Vector Machine (SVM) (Scheme and Englehart 2011; Atzori et al. 2014; Tavakoli et al. 2018; Wahid et al. 2018; Li et al. 2019; Kaczmarek, Mańkowski and Tomczyński 2019). It is known that the performance of HGR with the traditional approaches presented above will be greatly affected by the quality and quantity of the handcrafted features (Boostani and Moradi 2003; Phinyomark, Phukpattaranont and Limsakul 2012; Côté-Allard et al. 2020a). Therefore, the current focus has been on automated feature learning using deep learning techniques. Despite more recently proposed deep learning structures such as Long Short-Term Memory (LSTM) (He et al. 2018), Gated Recurrent Unit (GRU) (Nasri et al. 2019), Temporal Convolutional Network (TCN) (Bethauser et al. 2018; Zanghieri et al. 2020a), and Transformer (Montazerin et al. 2022), have been explored in this field, Convolutional Neural Network (CNN) remains the most widely used deep learning structure in feature learning and by far the most popular deep learning method for sEMG-based HGR (Phinyomark and Scheme 2018; Li, Shi and Yu 2021).

1.2.5 Post-processing

The post-processing techniques are usually simple but effective options, which can be applied to increase the robustness and accuracy of HGR regardless of the

methods used in each stage introduced above. The most commonly seen post-processing technique may be Majority Voting (MV) (Hudgins, Parker and Scott 1993; Amsüss et al. 2014; Geng et al. 2016; Du et al. 2017; Moin et al. 2018; Betthausen et al. 2020; Wahid, Tafreshi and Langari 2020), which only outputs a class that received the most votes based on consecutive predictions. Another intuitive approach that requires information from multiple consecutive windows is Bayesian Fusion (BF) (Khushaba et al. 2012). However, these methods can stabilise the prediction stream but introduce a response delay in the meanwhile. Note that although BF may only need two consecutive windows, a perceptible delay can be caused by disjoint windows, which is a requirement to comply with the independence condition assumed by Bayesian theory. A different category of post-processing technique is confidence-based rejection, which is used to improve the performance of HGR by not making uncertain predictions (Scheme, Hudgins and Englehart 2013; Scheme and Englehart 2015; Robertson, Englehart and Scheme 2019; Krasoulis, Vijayakumar and Nazarpour 2020; Wu et al. 2021; Bao et al. 2022). This method does not introduce additional delay but is highly dependent on determining the optimal rejection threshold.

1.2.6 Performance Evaluation

The offline performance of sEMG-based HGR is measured primarily on the basis of classification accuracy. This can be biased if the data for each class in the test set are not balanced. To avoid it, in addition to making the balanced test data set available, unbiased accuracy metrics can be used, such as macro-average accuracy, which weighs all classes equally (Hartwell, Kadiramanathan and Anderson 2018). Furthermore, more general evaluation metrics can be considered to fully measure the accuracy of a HGR task, as summarised by Sokolova and Lapalme (2009). From the perspective of the train test split, which refers to the training and test sets partition, recognition accuracy can be divided into several situations. When

considering the subject-specific performance, accuracy can be calculated in an intra-session or inter-session scenario, where each session comprises data collected by a subject performing several hand gestures in several repetitions on the same acquisition. Intra-session accuracy is computed when the training and test data are from the same session, whereas inter-session accuracy is computed when the training and test data are from different sessions. Similarly, inter-subject accuracy is usually computed to evaluate the performance of subject-independent HGR, which aims to construct a generic model trained by data collected from several subjects and tested for a new user.

Unlike offline testing, the performance of the complete myoelectric control system in real-time can be measured by evaluation metrics used to evaluate the target performance of a prediction model in the real world based on Fitts's law (Fitts 1992), including overshoot, throughout, path efficiency, completion rate, average speed, and stopping distance (Scheme and Englehart 2013; Scheme, Hudgins and Englehart 2013; Robertson, Englehart and Scheme 2019; Ameri et al. 2019).

1.3 Problem Statement

Despite promising results reported on sEMG-based HGR, especially with the aid of deep learning, a significant gap remains between current research results and clinical or commercial implementations. The reasons for this gap are summarised from the perspectives of practical considerations and technical issues.

1.3.1 Practical considerations

Current research in the field of sEMG-based HGR is often confined to improving the accuracy of offline recognition without considering the underlying issues that may arise for commercial applications. From the perspective of controlling devices such as prosthetic hands and hand exoskeleton robots based on myoelectric

sensing, the HGR system should be designed considering not only maximising recognition accuracy, but also minimising a series of costs, time spent on model training, delay, and energy consumption. One of the challenges is to consider the trade-off between these objectives. It is clear that recognition accuracy can be improved by introducing more electrodes, more extracted features, more complex models, or post-processing techniques that use information from more than one window. However, these will cause more costs for hardware and software used for data acquisition, data transfer, data storage, and data processing, together with additional model training time, latency, and possible control delay. To stay in focus and develop the system of sEMG-based HGR considering commercial viability, we have the initiative to face the challenges of HGR under more strict conditions, with only limited electrodes, no feature selection, simple architecture of deep learning models, and post-processing techniques without using information from more than one window.

Additionally, the black-box nature of machine learning or deep learning models may pose clinical challenges. For example, rehabilitation users may have reliability concerns about the myoelectric control system if recognition processing mechanisms cannot be easily interpreted from a safety perspective. Unfortunately, the model reliability in the field of sEMG-based HGR has not been adequately defined or investigated so far. Another clinical concern is due to the amount of training data available. Specifically, one cannot expect to collect a large amount of sEMG recordings from an end user in real life (Côté-Allard et al. 2019). Therefore, even though deep learning algorithms may work well with a large amount of training data, one of the clinical challenges comes from the fact that only limited data from an user are available. This is also one of the main factors that can cause low robustness of sEMG-based HGR in practical use.

1.3.2 Technical challenges

Despite taking the input of sEMG signals from forearm muscle groups is a natural and intuitive way for a prosthetic hand control system, the biggest challenge comes from the variability in the nature of sEMG signals, which is caused by intrinsic and extrinsic factors. Intrinsic factors include (i) the time-dependent, task-dependent, and stochastic nature of the neural drive to muscles (Rahimian et al. 2021); (ii) the alterations caused by muscle fatigue; (iii) the subject-dependent neural control strategies. It becomes more complicated when it comes to amputation, as many factors, such as the remaining percentage of the forearm and the sensation of the phantom limb, can also cause the variability of sEMG. Furthermore, the long-term robustness of sEMG-based HGR can be affected by several extrinsic factors, such as electrode displacement and skin conductivity (Hargrove, Englehart and Hudgins 2006; Kaufmann, Englehart and Platzner 2010; Scheme et al. 2010; Scheme and Englehart 2011; Young, Hargrove and Kuiken 2012). Note that even though the quality and quantity of sEMG signals can be improved by collecting from gel-based electrodes, its inconvenience may restrict practical application, as users are usually required to shave and wash the skin for optimal skin conductivity. For practical purposes, the sEMG signals used in this thesis are mainly collected from dry electrodes, thus highlighting the challenge of improving the robustness of sEMG-based HGR.

1.4 Aims and Objectives

For a noninvasive myoelectric sensing-based HCI system, sEMG-based HGR determines the upper limit of the performance of gesture-based interaction and object manipulation. This thesis focused on the problem of classifying hand movements from sEMG with a main view towards myoelectric control applications, but presents ideas relevant to other HCI applications as well. Therefore, the aim

of this thesis is to improve the robustness of sEMG-based HGR under the strict condition of considering practical needs for commercial applications by developing a set of methodologies and techniques to address the problems and challenges described in Section 1.3. Consequently, the objectives are summarised below.

- Investigate and compare various models capable of automatic feature learning from sparse multichannel sEMG signals. The goal is therefore to design end-to-end classifiers that directly take a limited amount of raw sEMG data as input.
- Integrate deep learning techniques with improved interpretability to enhance the current state-of-the-art (SoA) in understanding the Hand Gesture Recognition (HGR) process, aiming to explore and mitigate the black box concern associated with current approaches.
- Establish a systematic evaluation framework that provides fair comparison metrics to help design accurate, robust, and reliable sEMG-based hand gesture classifiers.
- Develop a solution to enhance the robustness of sEMG-based HGR for long-term usage, with a focus on minimising deterioration of inter-day performance.
- Propose a method to improve the performance of sEMG-based HGR using only limited data collected from the user.

1.5 Contributions

All the above practical considerations lead this thesis to address the challenges of HGR with a limited number of sparse multichannel sEMG signals. The key contributions of this thesis can be summarised as follows:

- Proposing a framework to design a simple but effective uncertainty-aware sEMG-based hand gesture classifier, that is, Evidential Convolutional Neural Network (ECNN), which can make its inference along with explainable quantified multidimensional uncertainties, given in Chapter 3, publications (Lin et al. 2022, 2023). This can provide better model interpretability, which mitigates the reliability concern of HGR directly.
- Providing reliability analysis for sEMG-based HGR by introducing a proper definition of model reliability, as shown in Chapter 4, publication (Lin et al. 2023). This fills the gap of lacking consensus on the definition of model reliability in this field and offers a new dimension of evaluation, which can help model selection together with recognition accuracy.
- Improving the long-term robustness of sEMG-based hand grasp recognition with proposed ECNN by rejecting uncertain predictions, as shown in Chapter 5, publication (Lin et al. 2023). This presents a practical solution to address the variability of sparse multichannel sEMG signals with almost no extra costs.
- Proposing a normalisation approach to leverage large amounts of inter-subject data for model training so that improving recognition accuracy of sEMG-based HGR, as shown in Chapter 6, publication (Lin et al. 2020). This practical solution addresses the clinical challenge of having only a limited amount of data from any individual.

To sum up, what we want readers to take away from this work is not just a specific model, but a theoretical framework to design reliable models for sEMG-based HGR.

1.6 Publications

This section shows the list of articles published in leading international conferences or journals in this field. The work in this thesis has resulted in the following papers:

(Journal Articles)

- Lin, Y., Palaniappan, R., De Wilde, P. and Li, L. (2022). Reliability analysis for finger movement recognition with raw electromyographic signal by evidential convolutional networks. *IEEE Trans Neural Sys & Rehab Eng*, 30, pp. 96–107
- Lin, Y., Palaniappan, R., De Wilde, P. and Li, L. (2023). Robust long-term hand grasp recognition with raw electromyographic signals using multidimensional uncertainty-aware models. *IEEE Trans Neural Sys & Rehab Eng*, 31, pp. 962–971

(Conference Proceedings)

- Lin, Y., De Wilde, P., Palaniappan, R. and Li, L. (2018). Muscle connectivity analysis for hand gesture recognition via sEMG. In *2018 Signal Inf Process Assoc Annu Summit Conf APSIPA Asia Pac. (APSIPA ASC)*, IEEE, pp. 848–852
- Lin, Y., Palaniappan, R., De Wilde, P. and Li, L. (2020). A normalisation approach improves the performance of inter-subject sEMG-based hand gesture recognition with a ConvNet. In *EMBC*, IEEE, pp. 649–652

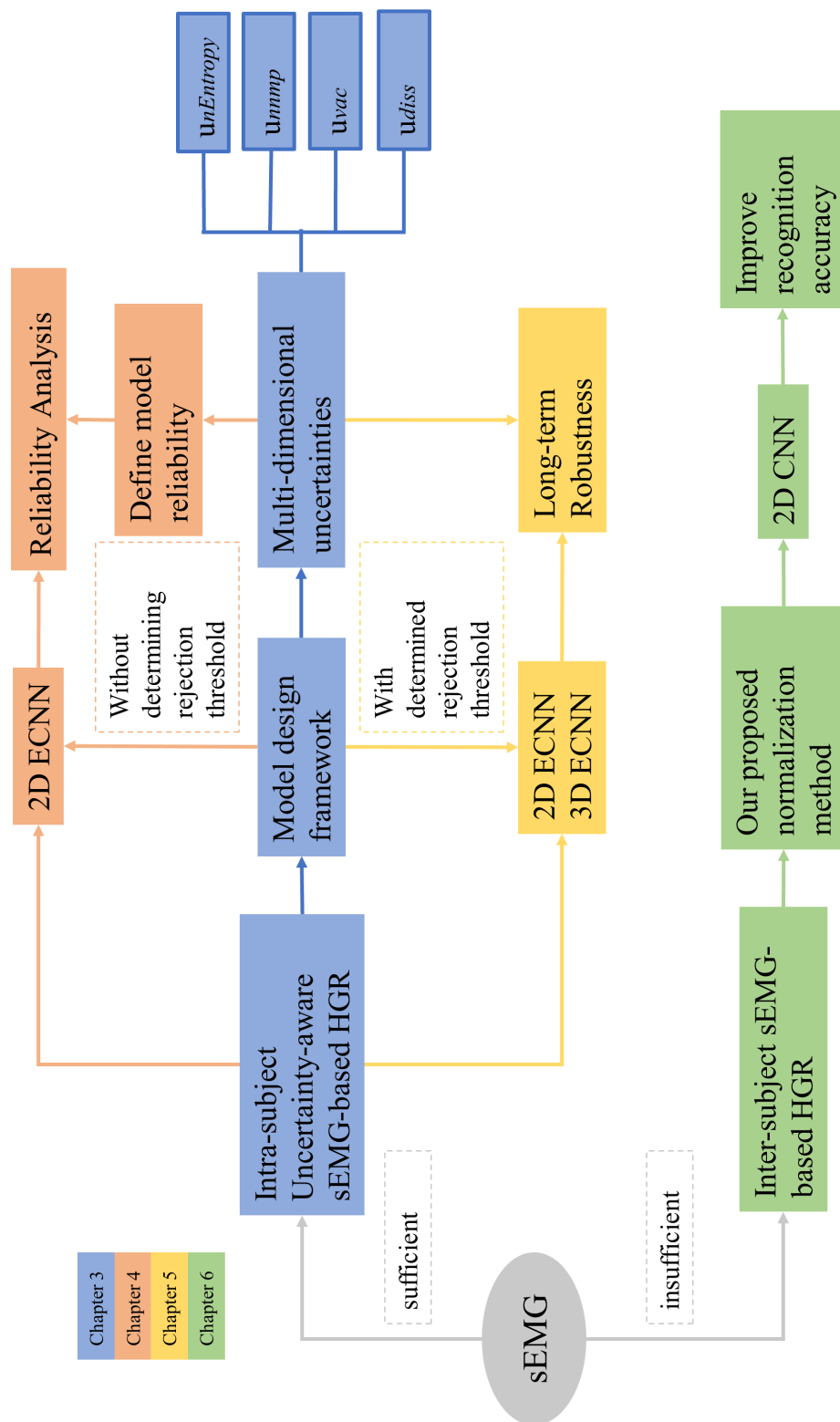


Figure 2: The high level structure of this thesis.

1.7 Thesis Structure

The structure of this thesis is organised as follows:

Chapter 2 provides the necessary background information and a systematic literature review for analysing state-of-the-art hand gesture recognition (HGR) using sparse multichannel surface electromyography (sEMG) signals and deep learning.

Chapter 3 addresses the following fundamental questions: 1. What does uncertainty awareness mean in the context of sEMG-based hand gesture models? 2. Why is considering uncertainty crucial and worth contemplating in this domain? 3. How do we design uncertainty-aware sEMG-based hand gesture models?

Chapter 4 then addresses the following two key questions: 1. How to evaluate the ability of an uncertainty-aware model in a fair and comparable way? 2. How this can reform the current evaluation system of sEMG-based HGR.

Furthermore, the exploration of how to effectively leverage the uncertainty inferred by uncertainty-aware models in practical use is presented in Chapters 4 and 5.

In Chapter 6, the subject-independent sEMG-based HGR is investigated with a proposed normalisation method.

Finally, Chapter 7 concludes the thesis by providing a systematic summary of the entire work and identifying potential directions for future research in this field as an extension of this thesis.

Chapter 2

Literature Review

This chapter presents the necessary background information and a systematic review of the literature to analyse state-of-the-art (SoA) Hand Gesture Recognition (HGR) using sparse multichannel Surface Electromyography (sEMG) signals and deep learning. The key topics covered here are the fundamental description of Electromyography (EMG) and the literature review of sEMG-based HGR. The main goal is to identify recent trends and research gaps in the field of sEMG-based HGR.

2.1 Electromyography

2.1.1 Discovery History

As shown in Figure 3, the era of bioelectricity began in the late 18th century, with the first empirical evidence of energy production in the form of bioelectricity approved by *Luigi Galvani* and the time course of the action potential was first recorded by *Julius Bernstein*, with the help of *Emil du Bois-Reymond* (Schuetze 1983). While the possibility of recording electrical activity during voluntary muscle contraction has been discovered, *Étienne-Jules Marey* made the first recording of electrical muscle activity during voluntary contractions in 1890 and introduced

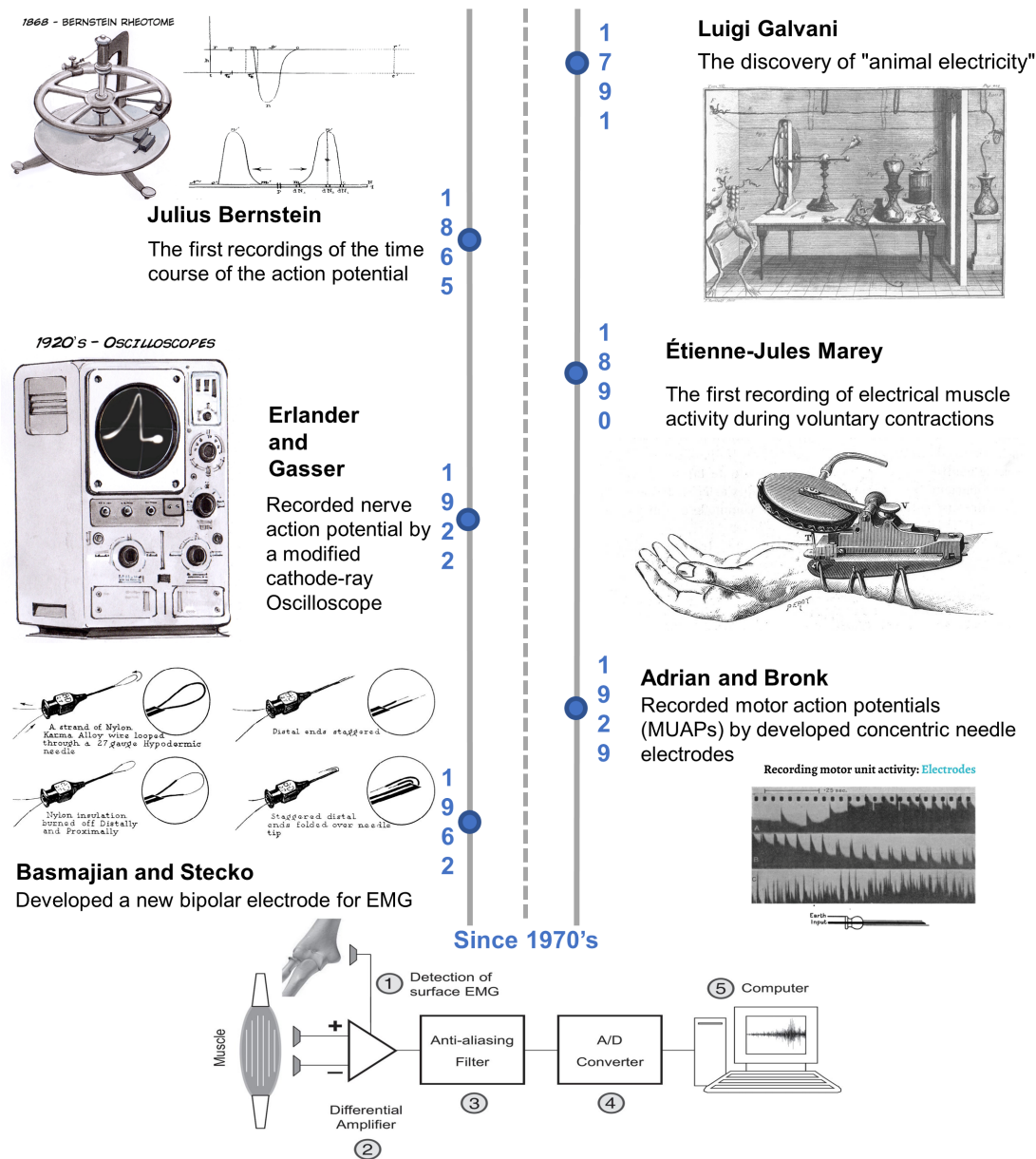


Figure 3: A brief history of the discovery of Electromyography.

the term EMG (Kleissen et al. 1998; Chakrabarti et al. 2015). Since Gasser and Erlanger had shown that nerve action currents can be easily and accurately recorded with an oscilloscope in 1922, more advanced technology has been developed with the main objective of improving the ability to collect electrical signals from muscles with the little invention. In 1929, Adrian and Bronk (1929) had successfully

designed and developed concentric needle electrodes to measure Motor Unit Action Potentials (MUAPs) from a small volume of muscular tissue within deep muscles. Such electrodes are still active for diagnostic neuro EMG today (Blanc and Dimanico 2010). An alternative approach to collect EMG has been developed by Basmajian and Stecko in 1962 to use fine-wire electrodes to record muscle activity from deeper and smaller muscles. Since 1970s, thanks to the development of modern computers and a number of EMG collection techniques, EMG signals have become an effective tool in the field of clinical diagnostics and motor control. For example, Herberts et al. had employed sophisticated digital computer techniques to improve hand prosthesis control using myoelectric patterns as early as 1973.

2.1.2 Biological Basis

Generally speaking, the EMG signal measures the electrical activity produced by the skeletal muscles, which refers to the muscle where its tissue is attached to the bone and its contraction is responsible for the support and movement of the skeleton. Electrical activity is generated directly by the central nervous system during each muscle contraction and is highly dependent on the anatomical and physiological properties of the muscles. Specifically, as shown in Figure 4, muscle fibres produce muscular force by receiving an electrical impulse that is propagated from the motoneuron to each motor endplate once the central nervous system activates this motoneuron.

From the perspective of the neuromuscular system, the EMG signal measured by the electrodes located in muscle mass during each voluntary or electrically elicited contraction is comprised of the linear summation of MUAPs from all active Motor Units (MUs) (Lowery 2009). Note that MU represents the anatomical and functional element of the neuromuscular system and is formed by the alpha spinal

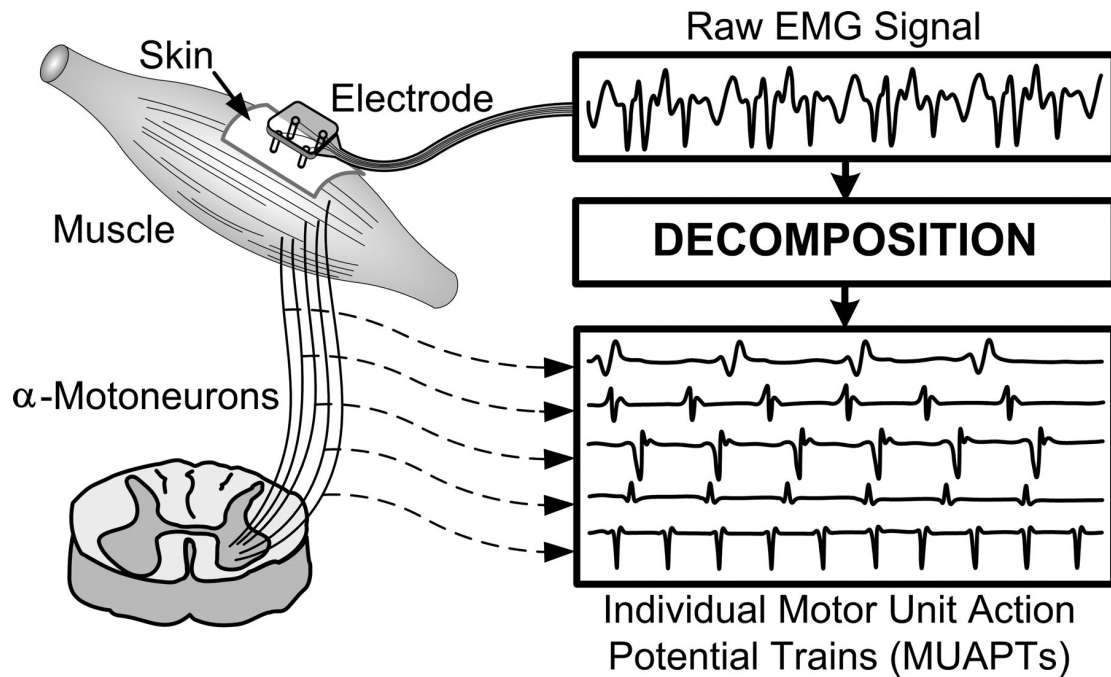


Figure 4: The outline of the decomposition of the sEMG signal into its constituent MUAPs. (Luca et al. 2006)

motor neurons and its innervated set of muscular cells (Rodríguez-Carreño, Gila-Useros and Malanda-Trigueros 2012). In fact, what EMG signal records are the electrical changes generated by the activity of MU, which is called MUAP. The illustration of how the action potential of each fibre contributes to the EMG signal can be seen in Figure 5. It clearly shows that the depth of the fibre is the main factor that affects the contribution that the action potential of each fibre makes to the EMG signal. Specifically, deeper muscle fibres contribute less to the sEMG signal (Kamen 2013). It can be seen that muscle fibres 3, 4, and 5 are farther away from electrodes than fibres 1 and 2 while actually contributing relatively smaller action potentials.

Each motoneuron is known to innervate a group of muscle fibres where the number varies from 10 to several thousand. The number of muscle fibres innervated by a motoneuron is termed the *innervation rate*. This is one of the main factors that can determine the amplitude of the MUAP. Generally, a larger MUAP

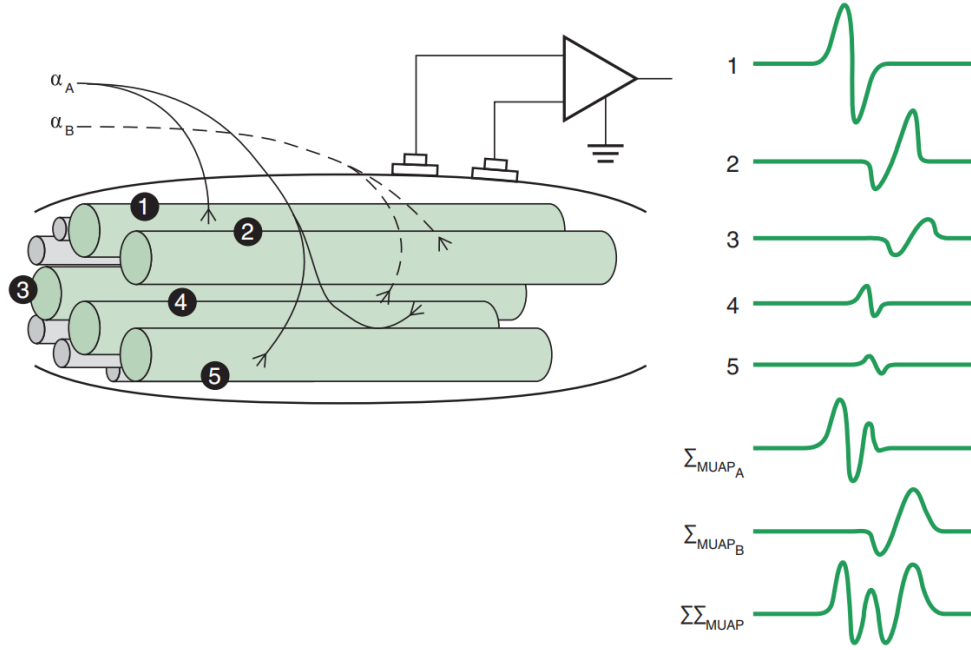


Figure 5: The illustration of how the action potential of each fiber contributes to the EMG signal. (Kamen 2013) There are 5 muscle fibers in this example as labelled in the left panel while their corresponding action potentials are shown in the right one. Furthermore, two MUs, α_A and α_B , are shown here. The amplitude of each MUAP is represented as the algebraic sum of action potentials of the contributed muscle fibers. Eventually, the recorded EMG signal is the algebraic sum of all MUs.

is expected to be observed with a higher innervation rate. Another one is *firing rate*, which is a term used to represent how frequently motor units are activated by the nervous system. To better understand this process, a linear model of EMG is presented as follows:

$$EMG(t) = \sum_{i=0}^{N-1} H_i * U_i(t) + w(t) \quad (1)$$

where N is the total number of MUs, H_i and U_i represent the MUAP and the spike train or firing impulse of i -th MU, respectively, $*$ denotes convolution, w accounts for noise, and the zero mean Additive White Gaussian Noise (AWGN)

is commonly used for modelling. This model can be simply visualised in Figure 6. Additionally, it is worth mentioning that the muscular force of a single muscle

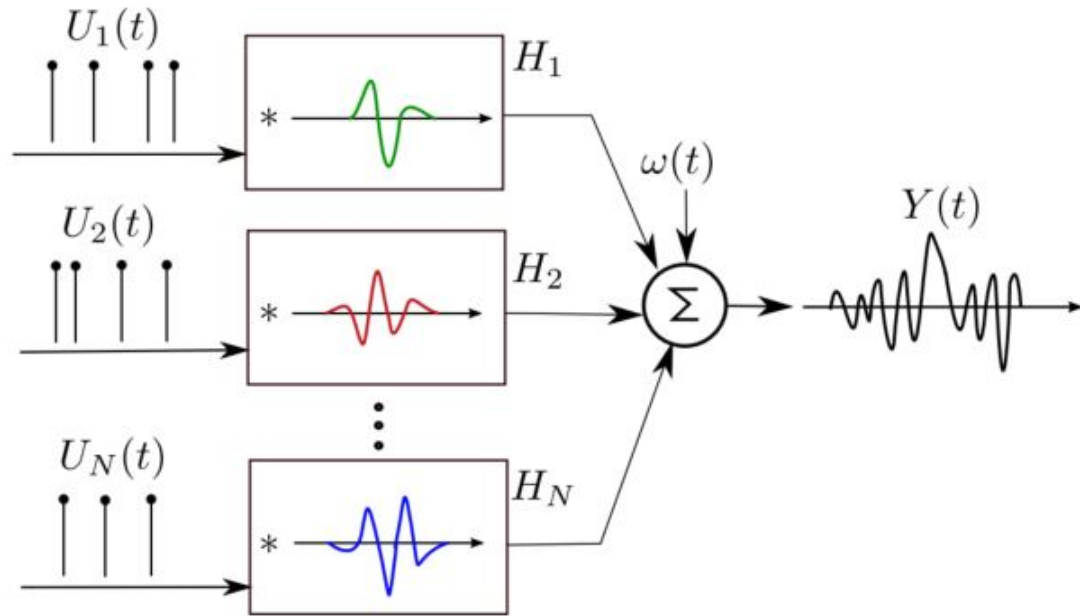


Figure 6: The illustration of mathematical model of EMG signal. (Akhmadeev 2019)

is initiated by activating an increasing number of motor units. Moreover, during the process of performing any muscular action, initially, smaller motor units are recruited first, and then, as the force requirement increases, the larger motor units are recruited successively. Therefore, it can be assumed that there is a positive relationship between the firing rate and the amount of muscular force.

As a result, the amplitude of the signal EMG can reflect the extent of muscular force. Since the EMG signal is a composite electrical sum of all active motor units, a large peak shown at a specific time point could therefore be the result of the activation of two or more motor units. In contrast, there are two possible reasons that can explain the situation when the signal crosses the baseline, which is often at zero volts. This could be either the number of activated motor units is low or the result of the balance caused by positive and negative phases of different

MUAPs. Under normal circumstances, it can be assumed that the amplitude of EMG increases as the intensity of muscular contraction increases. Strictly speaking, this is true only under an isometric condition, which is a condition of muscle contraction where muscle tension changes but muscle length remain the same (Widmaier, Raff and Strang 2010). For example, if a participant is required to hold a dumbbell in a fixed position, incremental changes related to the amplitude of the sEMG signal collected by an electrode attached to his or her biceps would be observed when incremental changes in muscular force are produced.

Understanding the relationship between EMG and muscular force is an important research topic, which has direct benefits for applications of electromyographic techniques. Taking the control of prosthetic limbs as an example, this can provide a guide on how much electrical activation should be used to achieve the desired level of force. To gain a better understanding of the EMG-force relationship, many factors need to be considered, such as the type of muscle contraction, the size of the muscles involved, and the potential role of various agonists and antagonists. Currently, most of the literature has investigated only the relationship between EMG activity and external force production, which mainly represents the action of objects that contact the body. Correspondingly, a full understanding of the relationship between muscle electrical activation and internal forces, which usually refer to responses of our body to external forces, is lacking.

2.1.3 Influence factors

To avoid misuse or misinterpretation of the EMG signal, it is necessary to know the myriad of sources that can influence the signal. Knowing these can also help us to better record the EMG signal and optimise the signal analysis algorithm (Reaz, Hussain and Mohd-Yasin 2006). Briefly, factors that affect the EMG signal can be categorised as either intrinsic or extrinsic.

Intrinsic Factors

Intrinsic factors include any influences from physiological, anatomical, or biochemical perspectives. In addition to the previously introduced innervation rate and firing rate, intrinsic influences include blood flow, muscle length and depth, the thickness of the subcutaneous fat layer, and the tissue between the muscle and surface electrodes. Both tissue and increased muscle blood flow have a low-pass filtering effect, causing the result that the high-frequency characteristics of the action potentials of deeper muscle fibres are attenuated appreciably at the surface (Kamen 2013). Note that blood flow increases dramatically during muscle contraction, which also leads to changes in muscle length. The amplitude of the muscle action potential decreases as its length increases (Gerilovsky, Tsvetinov and Trenkova 1986; Hashimoto et al. 1994).

Furthermore, as stated by Roeleveld et al., EMG amplitude decreases as the distance between the surface electrodes and the muscle increases. Consequently, the quality of the collected EMG can be affected by the thickness of the subcutaneous fat layer. Higher levels of subcutaneous fat are commonly associated with lower Signal-to-Noise Ratio (SNR). It is worth mentioning that this is one of the most significant factors that can explain more than 50% of the variance in the EMG signal (Nordander et al. 2003). There are many other underlying intrinsic factors that can affect EMG signals, such as muscle fatigue, dehydration, or interruption in muscle blood flow. All of these factors together are the reasons why EMG signals are difficult to interpret properly, especially during dynamic contractions.

Extrinsic Factors

Extrinsic factors refer mainly to signal recording equipment, especially sEMG sensors or electrodes. Specifically, both the amplitude and frequency of the EMG signal can be affected by electrode characteristics, including the material used for

conducting surfaces such as silver, gold, and stainless steel, the electrode size, the electrode placement, and even the interelectrode distance. As defined in the SENIAM¹, electrode size is defined as the size of the conductive area of a sEMG electrode. This is directly proportional to the amplitude of the detected sEMG signal. However, the electrode becomes unwieldy as its size increases. Furthermore, an electrode with a larger size cannot be used to detect EMG signal from relatively small muscles (Luca 2002). Therefore, a compromise often needs to be established between increasing the signal amplitude and reducing the electrode size.

Undoubtedly, electrode placement has the most straightforward effect on the quality of the recorded EMG signal. As can be seen in Figure 4, the preferred electrode location is between the motor point and along the longitudinal midline of the muscle, to cover as many muscle fibres as possible. With regard to the interelectrode distance, Mezzarane et al. pointed out that the selectivity of the surface electrodes decreases as the electrodes are closer to each other. More importantly, both the proper design of the electrodes and the carefully selected placement of the electrode can help reduce the effect of crosstalk, which means that the detected EMG signal will record muscle activity not only from the target muscle but also neighbouring ones. Cross-talk cannot be completely avoided, which is one of the most important reasons why sEMG is more useful when the global activity of the muscle is of interest.

In addition, the electrode-skin interface can affect the EMG signal. In detail, a poor electrode-skin interface can lead to high electrode impedance, resulting in poor biological signal quality, that is, a poorer SNR (Yamamoto and Yamamoto 1978; Neuman 2010). There are many factors that can affect the electrode-skin interface, such as perspiration (Abdoli-Eramaki et al. 2012), humidity, and electrode

¹The SENIAM project (Surface ElectroMyoGraphy for the Non-Invasive Assessment of Muscles) project is a European concerted action in the Biomedical Health and Research Program (BIOMED II) of the European Union.

shifts.

2.1.4 Characteristics of sEMG

To sum up, there are a few key characteristics of sEMG. From the perspective of the time domain, its amplitude is assumed to be stochastic in nature. Even though the range of this depends on the muscle types and conditions, the peak-to-peak amplitude can usually range from 0 to 10 mV (Luca 2002). Regarding the electromyographic frequency characteristics, the frequency spectrum of the sEMG signal is positively skewed (Kamen 2013). In general, the usable frequency component can range from 0 to 500 Hz, while the most dominant spectrum is in the 50-150 Hz range. Note that a signal is assumed to be stationary when computing its spectral frequency content. However, the EMG signal recorded during dynamic contractions violates the assumption of stationarity. This may be solved by introducing traditional approaches for analysing nonstationary signals, such as Short Time Fourier Transform (STFT) and Wavelet Transform (WT).

2.1.5 Applications

Electromyography signals have been widely used for clinical diagnosis and Human-Computer Interaction (HCI) applications due to valuable information on the nerve system they carry.

Clinical Diagnosis

For clinical applications, EMG signals are often used as a diagnostic tool to detect muscular disorders, gait disorders, and seizures by recognising abnormal patterns. As noted in a recent review (Papagiannis et al. 2019), gait analysis using sEMG with kinematic and kinetic data is a significant diagnostic procedure for patients who suffer from musculoskeletal disorders. Furthermore, as stated by Beniczky,

Conradsen and Wolf, a specific seizure, i.e., bilateral tonic-clonic seizure, can be detected by measuring the quantitative sEMG changes in the high-frequency component. However, it should be noted that the needle EMG can be used more often than sEMG for the purpose of clinical diagnosis, especially when the neurophysiological characteristics of neuromuscular diseases need to be evaluated. This is because spontaneous activity occurring at the single fibre level, such as fibrillations, positive spikes, and myotonia, is difficult to measure from only the skin surface of each individual (Drost et al. 2006).

Human-computer interaction applications

In contrast, sEMG has been widely exploited in the field of HCI in recent decades owing to its non-invasive property and the development of sEMG data acquisition. For example, a low-cost portable device, Myo armband (Thalmic Labs), has recently been proposed as a commercial sEMG acquisition that can send data to computers easily by wireless communication. As can be seen in Figure 7, Myo armband is a wearable device with eight EMG sensors and can be used to identify five hand gestures by internal algorithms. It is no exaggeration to say that this significantly promotes research on sEMG-based HCI, as evidenced in a large number of papers since 2014 (Samadani and Kulić 2014; Kutafina, Laukamp and Jonas 2015; Morais et al. 2016; Kerber, Puhl and Krüger 2017; Pizzolato et al. 2017; Zia ur Rehman et al. 2018a; Wahid et al. 2018; Hartwell, Kadiramanathan and Anderson 2018; Sun et al. 2019; Prahm et al. 2019; Côté-Allard et al. 2019; Hu et al. 2019; Betthauser et al. 2020; Fatimah et al. 2021).

Within the context of HCI, EMG signals, usually collected from the forearm muscles, are regarded as a source of system input for either healthy subjects or users with physical disabilities to the functionality offered by computer systems. More specifically, such HCI systems, which are controlled by transforming myoelectric signals into interaction commands that convey the intent of the user's

movement, can be referred to as Muscle-Computer Interaction (MCI) (Saponas et al. 2008; Chowdhury, Ramadas and Karmakar 2013; Geng et al. 2016; Du et al. 2017; Li et al. 2019; Wei et al. 2019).



Figure 7: The illustration of Myo armband sensor by Thalmic labs, which has ceased production recently and no corresponding website is available now. The bottom panel has shown all five possible hand gestures that can be classified from built-in programs (Tatarian et al. 2018).

Improving the efficiency of communication is one of the main purposes of using MCI. For example, Sign Language Recognition (SLR), which helps translate sign language into text or speech, can be implemented by employing sEMG sensors placed on the forearm. Compared with conventional approaches for SLR such as employing a camera, this provides a wearable solution with both low-cost and high-portable to overcome the communication barriers existing between the deaf and hearers who are incompetent in sign language (Cheng et al. 2015; Abreu et al. 2016; Savur and Sahin 2016; Yang et al. 2017b).

Additionally, another primary purpose of using MCI is to control a device

with the goal of rehabilitation in most cases, which is often also referred to as myoelectric control. One specific example of it is to assist people with compromised movement abilities by controlling exoskeletons, which are wearable robots that exhibit close physical and cognitive interaction with users, to amplify their movements so that normal functions can be restored (Khokhar, Xiao and Menon 2010; Leonardis et al. 2015; Lu et al. 2017; Trigili et al. 2019).

Another important application is to control a robotic prosthetic hand, especially for amputees who lose their limbs (Boostani and Moradi 2003; Khushaba et al. 2012; Atzori and Müller 2015; Krasoulis et al. 2017; Beaulieu et al. 2017; Tatarian et al. 2018; Akhmadeev 2019; Cognolato et al. 2020). Using sEMG recorded from the remnant muscles of the stump can be considered the most natural way for them to perform activities of daily living with the aid of prostheses. This is the reason why sEMG is virtually the only signal of control for the practical use of multifunctional prosthetic prostheses with an upper limb since the 1950s despite other signals such as Electroencephalogram (EEG) can also theoretically be the choice (Jiang et al. 2012b).

A classic and conventional technique to activate a function of the myoelectric-controlled prosthesis is to compare the sEMG amplitude to a predefined threshold. This simple approach has been proposed more than 60 years ago and is still used by the vast majority of commercially available powered prostheses. However, by employing this on-off control approach, the prosthetic hand is always moving at a constant speed (Moqadam et al. 2018; Mohebbian et al. 2021). To improve it, proportional myoelectric control is raised to allow amputees control over speed and force of grip by linking the mechanical output such as force, velocity, and position with sEMG signals continuously (Sears and Shaperman 1991; Fougner et al. 2012; Mohebbian et al. 2021). However, the main disadvantage of either on-off or proportional control, is that controlled movements are limited to the number of sEMG channels. To further improve the dexterity of myoelectric prosthetic

devices and overcome the limitations of conventional control approaches, pattern recognition-based control is recently developed based on feature engineering and classification techniques. The core of pattern recognition-based control is the sEMG-based HGR, which will be introduced in the next section.

2.2 sEMG-based Hand Gesture Recognition

In recent years, the research directions of sEMG-based HGR have undergone significant shifts in three key aspects. Firstly, the focus has shifted from identifying static hand gestures to recognising dynamic gestures. Secondly, the dominant method has moved from traditional machine learning algorithms to more sophisticated deep learning techniques. Lastly, the primary goal has shifted from merely improving accuracy to achieving robustness in recognition performance. These changes reflect the ongoing efforts to advance the SoA in sEMG-based HGR and are a testament to the ongoing development and growth of this field.

2.2.1 Evolution from static to dynamic HGR

Hand movements can be categorised into dynamic and static types, based on the type of skeletal muscle contraction involved (Jaramillo-Yáñez, Benalcázar and Mena-Maldonado 2020; Li, Shi and Yu 2021). Dynamic hand movements are executed through isotonic contractions, which involve changes in muscle length. Isotonic contractions generate force and movement, as the muscle shortens or lengthens to produce motion. On the other hand, static postures are maintained through isometric contractions, which do not involve any change in muscle length. Isometric contractions generate force without movement, as the muscle remains in a fixed position. When it comes to sEMG signals, there are also two states: steady and transient. Briefly, sEMG signals in the steady state are generated when

a hand gesture is sustained, while sEMG signals in the transient state are generated during the transition between gestures (Hudgins, Parker and Scott 1993). Therefore, the static HGR classifier uses only steady-state sEMG signals, while the dynamic HGR classifier utilises a combination of transient and steady-state sEMG signals.

It is evident that recognising dynamic hand gestures using sEMG signals is more challenging than recognising static hand gestures. This is because the transient part of the sEMG signal in dynamic gestures introduces more variability, making the gesture recognition process more complex. Therefore, static HGR used to be the mainstream, particularly when it comes to machine learning algorithms. Jaramillo-Yáñez, Benalcázar and Mena-Maldonado reviewed 65 studies that utilised sEMG and machine learning for HGR and found that only one attempted to recognise dynamic hand gestures. Furthermore, the HGR using sEMG data in the steady state is more accurate than in the transient state. For example, in as early as 2001, Englehart, Hudgins and Parker had already demonstrated remarkable recognition accuracy of 99.5% for four hand classes and 98% for six motion classes using only four channels of steady-state sEMG signals with the Linear Discriminant Analysis (LDA) classifier.

Although the classifier trained with steady-state sEMG signals performs well, it may not be suitable for dealing with transient sEMG signals during muscle contractions. This can result in unwanted mechanical vibrations during the transient state between movements, ultimately reducing the stability and lifetime of the mechanical system (Yang et al. 2012). Moreover, dynamic hand movements are considerably more advantageous than static ones, particularly in regards to performing Activities of Daily Livings (ADLs). Consequently, with the progress of deep learning, which is more adept at handling transient-state analysis (Li, Shi and Yu 2021), an increasing number of studies have recently turned their attention to investigating dynamic HGR over the past decade (Scheme and Englehart

2011; Chen et al. 2017; Phinyomark and Scheme 2018; Côté-Allard et al. 2020b; Li, Shi and Yu 2021).

2.2.2 Advancement from feature engineering to feature learning

Feature engineering refers to a set of techniques that are based on the manual creation or selection of features, which are commonly used to represent data in a machine learning system. It has traditionally been the main approach for sEMG-based HGR, where hand-crafted features are designed and selected by researchers. However, in recent years, feature learning has emerged as a powerful alternative, with deep learning techniques that demonstrate superior recognition performance compared to carefully designed hand-crafted features.

Feature learning involves automatically learning representations of the data through a hierarchy of layers in a neural network. In the case of sEMG-based HGR, this involves learning a set of features from the raw sEMG signals that are optimised for the specific HGR task. As the field of sEMG-based HGR moves to the era of “big data”, the paradigm is now rapidly shifting from feature engineering to feature learning (Phinyomark and Scheme 2018).

The sEMG-based HGR is often treated as an image classification problem, where each sEMG image is a frame of multichannel sEMG signals segmented by a sliding window. As such, Convolutional Neural Networks (CNNs) have been well investigated in the past decade and have achieved promising performance regardless of testing conditions on many public benchmark datasets. Recently, Temporal Convolutional Networks (TCNs) (Bai, Kolter and Koltun 2018) have been found to substantially outperform typical deep neural network models for modelling sequence data, such as Long Short-Term Memorys (LSTMs) and Gated Recurrent Units (GRUs), in a wide range of tasks and datasets. Compared to Recurrent Neural Network (RNN) models, the training process of a TCNs does

not require backpropagation at each sequential input. Such a highly parallelisable and light architecture allows TCN to mitigate the issues of gradient vanishing and explosion and requires less memory to speed up training.

Taking sEMG decoding as a sequential modelling problem, Tsinganos et al. first designed a TCN to address sEMG-based HGR and achieved an accuracy of as high as 89.76% in classifying 53 movements. However, this promising result was obtained by using only complete sequences instead of a sliding window ($\leq 300ms$), which is not feasible for real-time implementation. Taking into account the temporal dependencies within the input sequences where each sequence is taken by a sliding window of 200ms, Betthausen et al. found that the proposed TCN produces a more accurate and stable prediction performance, especially during interclass transitions, compared to other five models including LSTM, k -nearest neighbor (kNN), Support Vector Machine (SVM), Random Forest (RF), and Artificial Neural Network (ANN). The model *stability* was measured by evaluating its class-switching behaviour relative to the ground truth one using the proposed quantifiable stability metric. They also argued that addressing sEMG decoding as a sequential modelling problem will lead to enhancements in the robustness, responsiveness, and movement complexity available in prosthesis control systems in their subsequent work (Betthausen et al. 2020).

Furthermore, Zanghieri et al. proposed a TCN topology (TEMPONet) to address the temporal variability of sEMG-based hand grasp recognition by testing both intrasession and intersession validation. To further validate the practical performance of this model, they also proposed a real-time embedded platform for sEMG-based HGR and tested the quantised TCN not only on classification accuracy but also other practical metrics such as execution time and energy consumption on each classification (Zanghieri et al. 2020a). As a result, TEMPONet achieved SoA performance in Non-Invasive Adaptive Hand Prosthetics (NinaPro) DB6, producing a classification accuracy in the intrasession and between sessions

of 54.2% and 49.4%, respectively. Note that NinaPro DB6 involves different but very similar grasps, and there exist strong transient states in the timing of class switching.

2.2.3 Transition from accuracy to robustness

Over the past few decades, academic research in the field of sEMG-based HGR has primarily concentrated on enhancing the accuracy of classification. However, this has resulted in a disparity between the practical use of myoelectric control systems and laboratory settings. As mentioned in Section 2.1.3, there are numerous factors that can impact sEMG signals, which can alter the extracted features and subsequently reduce the recognition performance over time. This highlights the critical issue of robustness in sEMG-based HGR, particularly for applications in myoelectric control (Jiang et al. 2012b).

To bridge this gap, researchers have shifted their focus towards studying the robustness of sEMG-based HGR. For example, several studies have highlighted that variations in arm positions can result in a decline in HGR performance (Scheme et al. 2010; Fougner et al. 2011; Jiang et al. 2012a). This is because the feature extracted from the sEMG signals differ when performing hand movements in arm positions that are distinct from those used to train the classifier.

To address this issue, one straightforward solution is to train a multiposition classifier using sEMG signals collected from as many arm positions as possible (Geng et al. 2012). This approach is expected to increase the robustness of sEMG-based HGR by enhancing the generalisation ability of the classifier. Another solution is to design a two-stage cascade classifier (Geng et al. 2012; Geng, Zhou and Li 2012). In the first stage, a classifier is trained to predict the arm position of the amputee. In the second stage, each classifier is trained with data collected in a specific arm position, and the classifier that best corresponds to the

predicted arm position is selected to identify and track the specific hand movements.

Both of these solutions have been further validated using a motion-test environment, where the comparison results have shown that they outperformed the standard classifier by improving the real-time performance across seven classes of movements in five different arm positions (Geng et al. 2017). Specifically, the two-stage cascade classifier and multiposition classifier achieved an 8.7% and 12.7% increase in motion completion rate, respectively. These findings demonstrate the potential of these solutions to enhance the robustness of sEMG-based HGR and improve its practicality in real-world settings.

In order to address the decline in recognition accuracy that occurs over time, adaptive learning strategies can be implemented. These strategies involve using incremental learning schemes to update classifiers whenever new data is collected. Additionally, instead of simply adding all new data to the training set, it may be more effective to select a new sample set carefully. For instance, *gu2018robust* proposed an adaptive learning method that employs selectively chosen samples with ground-truth labels, and demonstrated its effectiveness in mitigating recognition degradation over one day.

Maintaining the accuracy of HGR using sEMG is a challenging problem, even when assuming the arm position is fixed. It has been reported by several studies that the classification accuracy can vary significantly over time due to day-to-day variations (He et al. 2015a; Huang et al. 2017; Zia ur Rehman et al. 2018b,a). Recently, deep learning techniques have been employed to tackle this issue. For instance, Bao et al. proposed a CNN with a novel confidence estimation mechanism that improves the long-term robustness of recognition by rejecting uncertain predictions. The proposed approach achieved an inter-session between-day accuracy of 73.33% under the rejection scheme.

2.3 Summary

In conclusion, the comprehensive analysis of current research directions in sEMG-based HGR has revealed significant shifts in focus, methodology, and objectives within this rapidly evolving field. As highlighted in the Literature Review, the transition from static to dynamic gesture recognition signifies a response to the demand for more natural and nuanced interactions. Moreover, the adoption of sophisticated deep learning techniques showcases the recognition of their potential in tackling complex sEMG signal analysis tasks. More importantly, the paradigm shift from accuracy-oriented approaches to prioritising robustness addresses the practical challenges of real-world applications. The Literature Review serves as a foundation for the subsequent chapters in this thesis, where we endeavor to build upon these insights and address specific gaps in the field. The ensuing chapters will explore uncertainty-aware models, a systematic evaluation framework to investigate the model reliability, and long-term robustness analysis to contribute to the advancement of SoA sEMG-based HGR. By leveraging these evolving research directions, we aim to pave the way for more reliable, interpretable, and adaptable HGR systems, ultimately enhancing user experiences and enabling new possibilities in myoelectric control applications.

Chapter 3

Methods

This chapter presents the framework proposed to design a simple but effective uncertainty-aware Surface Electromyography (sEMG)-based hand gesture classifier. Applying it allows us to quickly build a model with the ability to make its inference along with explainable quantified multidimensional uncertainties. This addresses the black-box concern of the Hand Gesture Recognition (HGR) process directly with better model interpretability.

3.1 Introduction

The Convolutional Neural Network (CNN), as the most popular deep learning model for sEMG-based HGR, has been successful in achieving better recognition accuracies compared to other machine learning algorithms. However, it has no ability to make inferences along with an explainable uncertainty about it. In other words, it is not an uncertainty-aware model (Kaplan et al. 2018). To extend CNN to accommodate uncertainty-awareness, we propose Evidential Convolutional Neural Network (ECNN) that integrates CNN with Evidential Deep Learning (EDL) (Sensoy, Kaplan and Kandemir 2018), which was developed based on the theoretical framework of Subjective Logic (SL) (Jøsang 2016). This allows

ECNN to be explicitly trained so that having the ability to infer understandable multidimensional uncertainties include *vacuity* (Sensoy, Kaplan and Kandemir 2018), expressing whether there is sufficient evidence to support model predictions, and *dissonance*, presenting the uncertainty derived from conflicting evidence (Jøsang, Cho and Chen 2018).

3.2 Subjective Logic

In Dempster-Shafer Theory of Evidence (DST) (Dempster 2008), a *frame of discernment* Θ is defined as a finite set of mutually exclusive elements in a domain, where a subset of Θ is referred to as a hypothesis or proposition, and a *singleton* is used to represent it if the cardinality of this subset is equal to 1. The belief of a proposition could be quantified by *belief functions* based on available evidence, which allows us to not follow the additivity principle of probability theory strictly, thus providing an additional “dimension of uncertainty” to make ignorance explicit (Reineking 2014). Based upon DST’s notion of belief assignment over Θ , SL (Jøsang 2016) provides a structured approach to connect beliefs to Dirichlet distributions so that we can approximate second-order Bayesian reasoning in a computationally efficient way. The second-order uncertainty of a multi-class classifier is represented by a Dirichlet Probability Density Function (PDF) over a multinomial distribution, which refers to the first-order uncertainty representing the predicted class probabilities. In other words, the first-order logic can be incorporated into the higher layer reasoning so as to enrich the uncertainty representation with extra information from beliefs.

Let $Y = (Y_1, Y_2, \dots, Y_K)$ be a discrete variable in a domain \mathbb{Y} , representing the class label. For a multiclass classification problem, the number of classes $K = |\mathbb{Y}| > 2$. A multinomial opinion over Y in SL is then defined as an ordered triplet $w_Y = (\mathbf{b}_Y, u_Y, \mathbf{a}_Y)$ where

- \mathbf{b}_Y refers to a *belief mass distribution* over \mathbb{Y} ;
- u_Y is the *uncertainty mass* that expresses the vacuity of evidence, which decreases as more observations in terms of statistical events are found;
- \mathbf{a}_Y represents a *base rate distribution* over \mathbb{Y} , which is known as *prior probability* in classic Bayesian theory.

The projected probability distribution of a multinomial opinion in SL defined as follows (Jøsang 2016):

$$\mathbf{P}_Y = \mathbf{b}_Y + \mathbf{a}_Y u_Y. \quad (2)$$

SL demonstrates clearly that there is a specific bijective mapping between a multinomial opinion and a Dirichlet PDF over the same domain \mathbb{Y} . Before proceeding further, let us recall the definition of a Dirichlet PDF over the same discrete variable Y on domain \mathbb{Y} (Ng, Tian and Tang 2011):

$$Dir(\mathbf{p}_Y) = \frac{\Gamma\left(\sum_{j=1}^K \alpha_j\right)}{\prod_{j=1}^K \Gamma(\alpha_j)} \prod_{j=1}^K p_{Y_j}^{(\alpha_j-1)}, \quad (3)$$

where \mathbf{p}_Y represents the probability distribution for the discrete variable Y , such that each $p_{Y_j} \in (0, 1)$ and $\sum_{j=1}^K p_{Y_j} = 1$; $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_K)$ is a strength vector of positive-valued Dirichlet parameters; $\Gamma(\cdot)$ is the standard Gamma function. Since the Dirichlet distribution belongs to the exponential family, its conjugation property allows us to consider the Dirichlet parameter $\boldsymbol{\alpha}$ as prior and observation evidence. From the perspective of SL, each singleton can have an arbitrary additive base rate distribution \mathbf{a}_Y over the domain \mathbb{Y} rather than the default value $1/K$ if no additional information is provided (Jøsang 2016), as can be seen below:

$$\boldsymbol{\alpha} = \mathbf{r} + \mathbf{a}_Y W, \quad (4)$$

where \mathbf{r} is a vector of evidence over variable Y ; W is a constant expressing the

non-informative prior weight. The evidence representation of Dirichlet PDF can then be obtained by substituting the above equation into Equation (3) and the expected probability distribution over \mathbb{Y} is (Jøsang 2016):

$$\mathbb{E}_Y = \frac{\boldsymbol{\alpha}}{\sum \boldsymbol{\alpha}} = \frac{\mathbf{r}}{W + \sum \mathbf{r}} + \mathbf{a}_Y \frac{W}{W + \sum \mathbf{r}}. \quad (5)$$

Intuitively, to build such a bijective mapping, the projected probability distribution defined in Equation (2) is supposed to equal the expected probability distribution defined in Equation (5). More specifically, the observed evidence in Dirichlet PDF could be simply mapped to the belief mass distribution, inferring that:

$$\mathbf{b}_Y = \frac{\mathbf{r}}{W + \sum \mathbf{r}}, \quad u_Y = \frac{W}{W + \sum \mathbf{r}}. \quad (6)$$

Note that the total belief mass approaches 1 (or 0) while the uncertainty of vacuity reaches 0 (or 1), as the total evidence goes to infinity (or 0). These properties match the additivity requirement of a multinomial opinion over Y , that is,

$$\sum \mathbf{b}_Y + u_Y = 1. \quad (7)$$

3.3 Evidential Deep Learning

Based on the well-developed theoretical framework of SL, EDL was proposed to help extend a deep learning model to accommodate uncertainty-awareness by introducing explicitly training process (Sensoy, Kaplan and Kandemir 2018). In EDL, the term *evidence* \mathbf{e} has been defined as a measure of the amount of support collected from the extracted features in favour of an input sample to be classified into a certain class. Recall that a discrete variable $Y = (Y_1, \dots, Y_K)$ represents the class label for a K -classification problem. The non-informative prior weight W is equal to K since a uniform prior PDF is required when there is no observation.

Naturally, each element of the base rate vector \mathbf{a}_Y is equal to $1/K$ without any extra information. Therefore, one can easily compute the belief mass vector as follows:

$$\mathbf{b}_Y = \frac{\mathbf{e}_Y}{K + \sum \mathbf{e}_Y}. \quad (8)$$

It is noted that the denominator is referred to as *total evidence* S , as evidenced by the mathematical expression presented below:

$$K + \sum \mathbf{e}_Y = \sum (\mathbf{e}_Y + 1) = S. \quad (9)$$

Accordingly, the belief mass vector \mathbf{b}_Y is expressed as the normalisation of the observed evidence vector \mathbf{e}_Y . Furthermore, there is a mapping between the parameter vector of Dirichlet distribution and the evidence vector, as revealed by the following equation:

$$\boldsymbol{\alpha} = \mathbf{e}_Y + 1. \quad (10)$$

By substituting it into Equation (5), the predicted probability vector can be computed from the mean of the corresponding Dirichlet distribution, as

$$\mathbf{p}_Y = \frac{\mathbf{e}_Y + 1}{\sum (\mathbf{e}_Y + 1)} \quad (11)$$

In practical scenarios, each item of \mathbf{e}_Y has the lower limit of 0 and no upper limit. When there is no observation, that is, the total evidence is equal to 0, the predicted probability for each class is the same as $1/K$.

3.4 Evidential Convolutional Neural Network

EDL considers the output of a neural network as the evidence vector. Therefore, to satisfy the property of the evidence vector, i.e., the collected evidence for each class should be greater or equal to 0, EDL takes the non-negative output of a

neural network after replacing the *Softmax* layer with an activation layer such as *ReLU* as the evidence vector for the predicted Dirichlet distribution. This actually builds a bridge between deep learning models and the theoretical theory of SL. As such, ECNN can be simply designed by integrating a CNN (Côté-Allard et al. 2019) with EDL.

The design of ECNN represents an innovative application of EDL in the domain of sEMG-based HGR. Our contributions go beyond the mere design of ECNN architecture; instead, they primarily stem from the comprehensive framework established for the design and optimisation of ECNN models. It is clear that the performance of the model is profoundly influenced by the choice of loss functions and the effectiveness of hyperparameter determination. First and foremost, the proposed framework of designing ECNN allows us to investigate and explore various loss functions tailored to the specific requirements of sEMG-based HGR. This exploration is crucial as it sheds light on the diverse regularisation techniques and their impact on model performance. More details are presented in section 3.5. Moreover, our extensive investigation of hyperparameter optimisation plays a pivotal role in enhancing model performance and fine-tuning ECNNs for optimal results. For example, instead of solely relying on *ReLU* as the activation function, we treat it as a hyperparameter and incorporate other potential activation functions, thereby providing researchers with a more diverse toolkit to enhance model expressiveness and flexibility. More details can be found in section 3.6.

Overall, with the aid of this framework, ECNN can be trained to learn to collect the evidence that leads to a subjective opinion used to predict hand gestures with the support of explicit uncertainty estimates.

3.5 Loss functions

For a K -class supervised HGR problem, given a sample i and let \mathbf{y}_i be a one-hot encoding of the ground-truth class of it with $y_{ij} = 1$ and $y_{im} = 0$ for all $j \neq m$ where j and m are class labels. The evidence vector generated based on the learned features after taking the sample i as input to ECNN can be expressed as $f(\mathbf{x}_i|\Theta)$, where Θ refers to the model parameters of ECNN. Based on Equation (11), the predicted probability of sample i for j^{th} hand gesture is computed as

$$p_{ij} = \frac{f(x_{ij}|\Theta) + 1}{\sum(f(\mathbf{x}_i|\Theta) + K)}. \quad (12)$$

The sum-of-squares loss \mathcal{L}_{SSL} can be computed by using the Euclidean distance between the one-hot encoding \mathbf{y}_i and the predicted class probabilities \mathbf{p}_i , as shown below:

$$\mathcal{L}_{SSL}(\mathbf{e}_i, \mathbf{y}_i) = \frac{1}{K} \sum_{j=1}^K (y_{ij} - p_{ij})^2, \quad (13)$$

where y_{ij} represents the true probability that the sample i belongs to class j , and p_{ij} denotes the predicted probability assigned by the model for class that the sample i for class j . The sum is taken over all K classes, and then divided by K to compute the average squared difference between the one-hot encoding and the predicted class probabilities.

The potential of ECNN can be further explored by employing different learning strategies which are based on the sum-of-squares loss \mathcal{L}_{SSL} . For the sake of convenience, only three ECNN variants will be investigated and referred to as ECNN-A, ECNN-B, and ECNN-C, where each employs a specific variant of loss function for training. Assume that there are N training samples in total, ECNN-A is trained by minimising the mean of \mathcal{L}_{SSL} directly, as seen in Equation (14):

$$\mathcal{L}^{(A)} = \frac{1}{N} \sum_{i=1}^N \mathcal{L}_{SSL}(\mathbf{e}_i, \mathbf{y}_i). \quad (14)$$

Furthermore, ECNN-B and ECNN-C will be trained by minimising the loss function $\mathcal{L}^{(B)}$, and $\mathcal{L}^{(C)}$ respectively, as seen in Equations (15) and (16):

$$\mathcal{L}^{(B)} = \frac{1}{N} \sum_{i=1}^N (\mathcal{L}_{SSL}(\mathbf{e}_i, \mathbf{y}_i) + \lambda^{(B)} \mathbb{E}_{(\mathbf{x}_i, \mathbf{y}_i) \sim \mathcal{D}} [\text{KL} [\text{Dir}(\mathbf{p}_{-k}; \boldsymbol{\alpha}_{-k}) \parallel \text{Dir}(\mathbf{p}_{-k}; \mathbf{1})]]), \quad (15)$$

where $\lambda^{(B)} = \min(1.0, t/s)$; t is the current training epoch number and s is an annealing step used to control the annealing effect of the Kullback-Leibler (KL) divergence term, $\text{Dir}(\cdot)$ is the Dirichlet distribution, \mathbf{p}_{-k} stands the predictive probabilities for all wrong classes while $\boldsymbol{\alpha}_{-k}$ refers to the predicted Dirichlet parameters except the one that is attributed to the ground truth class k ;

$$\mathcal{L}^{(C)} = \frac{1}{N} \sum_{i=1}^N (\mathcal{L}_{SSL}(\mathbf{e}_i, \mathbf{y}_i) + \lambda^{(C)} \mathbb{E}_{(\mathbf{x}_i, \mathbf{y}_i) \sim \mathcal{D}} [\text{KL} [\text{Dir}(\mathbf{p}_{-k}; \boldsymbol{\alpha}_{-k}) \parallel \text{Dir}(\mathbf{p}_{-k}; \mathbf{1})]]), \quad (16)$$

where $\lambda^{(C)}$ is a constant and the others are the same as Equation (15).

It is clear that the sole distinction between $\mathcal{L}^{(A)}$ and the other two loss functions lies in the introducing of a regularisation term. Incorporating the KL divergence term in the loss functions $\mathcal{L}^{(B)}$ and $\mathcal{L}^{(C)}$ introduces a crucial regularisation mechanism in the model. This term ensures that the model not only focuses on improving overall training accuracy but also guards against overconfidence in its predictions. The objective is to encourage the predictive probability of the true label to be higher than other classes while minimising the predictive probabilities assigned to incorrect classes. In other words, by including the KL divergence term, the loss function guides the model towards more calibrated and reliable predictions, preventing the occurrence of extreme probabilities for incorrect classes. This regularisation mechanism plays a pivotal role in promoting well-calibrated confidence scores and enhancing the model's robustness, especially in scenarios with limited training data or imbalanced class distributions. The intuition behind

is to address the issue of overconfidence and ensure the model strikes an appropriate balance between accuracy and uncertainty in its predictions, thereby yielding more trustworthy results in sEMG-based HGR.

Furthermore, the distinction between $\mathcal{L}^{(B)}$ and $\mathcal{L}^{(C)}$ lies in the parameter λ , which governs the effect of the KL term. In $\mathcal{L}^{(B)}$, λ is designed to scale the KL divergence term dynamically. Specifically, it allows the KL term to exert a lesser influence during the early stages of training and gradually increase its impact as the model’s performance improves. Notably, $\lambda^{(B)}$ remains fixed at 1 once the training epoch reaches s . On the other hand, in $\mathcal{L}^{(C)}$, $\lambda^{(C)}$ is a constant hyperparameter set to maintain the effect of the KL at a consistent and reasonable level throughout the training process.

One of the pivotal contributions of this thesis is the thorough exploration and comparison of these loss functions, which provides a comprehensive guide on designing efficient ECNNs for sEMG-based HGR. By delving into the differences between $\mathcal{L}^{(A)}$, $\mathcal{L}^{(B)}$, and $\mathcal{L}^{(C)}$, we gain valuable insights into the implications of their respective regularisation techniques and their effects on model performance.

The investigation not only highlights the significance of the introduced regularisation terms but also demonstrates the influence of varying λ values in $\mathcal{L}^{(B)}$ and $\mathcal{L}^{(C)}$. The dynamic scaling of λ in $\mathcal{L}^{(B)}$ effectively adapts the strength of the KL divergence term as the model evolves through different training epochs, offering a balanced regularisation effect. On the other hand, $\mathcal{L}^{(C)}$ introduces a constant and reasonable λ value, ensuring consistent regularisation throughout training, suitable for scenarios where maintaining a steady level of KL regularisation is preferred.

Understanding the differences and relative merits of these loss functions empowers researchers and practitioners to make informed decisions when designing ECNNs for sEMG-based HGR tasks. The insights gained from this exploration serve as a roadmap for enhancing model performance, optimising hyperparameter

settings, and tailoring the loss function to the specific requirements of diverse applications. This contribution significantly advances the state-of-the-art (SoA) in sEMG-based HGR, enabling more accurate and robust models, ultimately fostering the development of innovative myoelectric control applications and Human-Computer Interaction (HCI) paradigms.

3.6 Hyperparameters

Model performance depends not only on model training but also on Hyperparameter Optimisation (HPO), where hyperparameters are the parameters that control the learning process and must be determined before finding the optimal model parameters. The most commonly seen hyperparameters include batch size and learning rate. Batch size refers to the total number of samples used for training to update the model parameters. The learning rate controls the amount of change in the direction towards minimising the loss function in response to the estimated error when updating the model parameters. As mentioned in section 3.4, the activation layer *ReLU* can be used to replace *Softmax* to generate nonnegative evidence vector. However, this is not the only option since its similar function can be achieved by using other activation layers, such as *SoftPlus*. As such, the activation layer can be considered as a hyperparameter when training ECNN, which is referred to as evidence function. Additionally, the annealing step s presented in Equation (15) is a hyperparameter for ECNN-B while the λ presented in Equation (16) is a hyperparameter for ECNN-C.

3.7 Evidential uncertainty

The power of ECNN comes from its ability to generate multidimensional evidential uncertainty, including *vacuity* and *dissonance*, which are referred to as

evidential uncertainty of ECNN. This can help us to understand the source of the uncertainty. Vacuity denotes uncertainty due to lack of evidence or knowledge and can be calculated based on Equations (8) and (9), see below:

$$u_{vac} = u_Y = \frac{K}{\sum \mathbf{e}_Y + K}, \quad (17)$$

where K is the number of classes and \mathbf{e}_Y is the vector of observed evidence for each class. Dissonance represents the uncertainty due to conflicting evidence, derived from a sufficient number of conflicting evidence by comparing each two singleton belief masses (Zhao et al. 2020):

$$u_{diss} = \sum_{j=1}^K \left(\frac{b_j \sum_{m=1, m \neq j}^K \mathbf{Bal}(b_j, b_m)}{\sum_{m=1, m \neq j}^K b_m} \right), \quad (18)$$

where K is the number of classes and $\mathbf{Bal}(b_j, b_m)$ represents the relative mass balance between a pair of belief masses b_j and b_m for the sample i , equals 0 when $b_j + b_m = 0$ and $1 - \frac{|b_j - b_m|}{b_j + b_m}$ otherwise.

3.8 Summary

In this chapter, the background and theory of EDL have been introduced and it has been shown how ECNN can be easily designed using EDL. In addition, the details of HPO, model training, and explainable multidimensional uncertainties were presented to illustrate the framework to design uncertainty-aware sEMG-based hand gesture classifier.

Chapter 4

Reliability Analysis of Finger Movement Recognition

4.1 Introduction

As introduced previously, Surface Electromyography (sEMG)-based Hand Gesture Recognition (HGR) is a practical application of sEMG that has found wide usage in advanced prostheses control (Rezazadeh et al. 2012; Atzori and Müller 2015) and other rehabilitation applications (Ghassemi et al. 2019). It is crucial that the development of such a classification-based control scheme is highly dependent on the accurate and robust hand gesture predictions of users. As a result, current research on sEMG-based HGR has focused on improving its accuracy (Shen et al. 2019; Côté-Allard et al. 2019; Wei et al. 2019) and robustness (Gu et al. 2018; Zia ur Rehman et al. 2018a; Côté-Allard et al. 2019; Lin et al. 2020) with recent deep learning techniques. Note that model robustness can be summarised as the ability to remain accurate in practical scenarios under many factors that can affect prediction performance, such as electrode shifts, sweating, limb posture

and force changes, and day-to-day variation (Hargrove, Englehart and Hudgins 2008; Young, Hargrove and Kuiken 2011; Farina et al. 2014; Khushaba et al. 2014; Hwang, Hahne and Müller 2017; Gu et al. 2018; Prahm et al. 2019). A special case of robustness is to address subject variability when considering user-independent sEMG-based HGR (Côté-Allard et al. 2019; Lin et al. 2020).

Recently, the rejection of hand movements based on uncertainty measures has shown good potential as a general practical solution to improve the usability of sEMG-based myoelectric control by boosting both the accuracy and robustness of HGR (Scheme and Englehart 2015; Robertson, Englehart and Scheme 2019; Wu et al. 2021). Ideally, most inaccurate ambiguous predictions could be rejected by introducing additional information, such as entropy or the normalised maximum probability of the predictive distribution, for the indication of the confidence level. The intuition behind this is to address the concern where the gesture recognition process is considered as a ‘black box’ for myoelectric control (Scheme and Englehart 2015). In this chapter, we provide a definition of *reliability* R of an sEMG-based hand gesture classifier as the quality of its uncertainty measures that produce confidence scores on the predictions of test samples. Its reliability analysis then refers to the evaluation of R . This is supported by the commonly held opinion that accurate and robust HGR is considered reliable (Gu et al. 2018; Zia ur Rehman et al. 2018a), and the statement that accurate uncertainty estimation is one of the essential factors for the reliable application of deep learning (Kopetzki et al. 2021).

Although deep learning models, particularly those based on Convolutional Neural Networks (CNNs), have achieved state-of-the-art (SoA) performance by reporting high sEMG-based HGR accuracy, the reliability analysis of CNNs in this field has remained unexplored, which has become a growing necessity due to the vulnerability of deep learning models reported recently (Szegedy et al. 2014;

Kurakin, Goodfellow and Bengio 2016; Su, Vargas and Sakurai 2019). The reliability analysis has direct benefits for current studies, including latent concerns about model reliability in rejection-based HGR. For example, Wu et al. recently proposed a metric-learning guided CNN to enhance the robustness of myoelectric control systems by effectively rejecting novel patterns, i.e., no new classes were included in the training. It is evident that there is a positive correlation between the defined reliability R and the performance of the rejection-capable sEMG-based HGR. This implies that quantifying R could provide a useful indication of model performance without suffering from the limitations of evaluating its rejection-capable recognition performance, such as introducing extra evaluation measures (e.g., accuracy-rejection curve (Nadeem, Zucker and Hanczar 2010), false activation error (Hargrove et al. 2010)) and highly relying on determining the optimal rejection threshold (Robertson, Englehart and Scheme 2019).

Additionally, current uncertainty measures used in sEMG-based HGR do not provide meaningful insight into the predictions. Recent studies in the field of predictive uncertainty estimation have shown that evidential neural networks (Sensoy, Kaplan and Kandemir 2018; Zhao et al. 2020) modelled with Dirichlet-based uncertainty (Kopetzki et al. 2021) are more efficient in explicitly measuring uncertainties such as vacuity and dissonance (Jøsang, Cho and Chen 2018) with almost no additional computational cost, unlike other approaches such as Bayesian neural networks (Blundell et al. 2015) or ensemble models (Lakshminarayanan, Pritzel and Blundell 2017). The potential of applying Evidential Convolutional Neural Network (ECNN) to the sEMG-based HGR will be further explored in this chapter.

The main goal of this chapter is to propose a framework to directly quantify R , with a specific focus on the reliability analysis of individuated finger movement recognition with raw sEMG. Such movements are highly complex and versatile (Scherberger 2009), naturally raising the real need for reliability analysis.

We first employ an existing end-to-end CNN model (Côté-Allard et al. 2019) and propose an uncertainty-aware model, that is, the ECNN by integrating it with evidential deep learning. As a pilot study towards the reliability analysis of sEMG-based finger movement classifiers, the discussion begins by illustrating how multidimensional uncertainties generated by ECNN, such as vacuity and dissonance, can be accurately quantified and utilised to recognize finger movements that are difficult to classify, as compared to CNN.

Furthermore, a brief comparison of the performance of rejection-capable finger movement recognition between CNN and ECNN is provided as empirical evidence to support the intuition behind this research. Finally, and most importantly, we first recommend using a threshold-free evaluation metric called normalised Area Under Precision-Recall Curve (nAUPRC) (Boyd et al. 2012) to evaluate the task of misclassification detection, which is introduced to quantify R , to avoid the pitfall that current related evaluation metrics such as Area Under Receiver Operating Characteristic (AUROC) (Fawcett 2006) and Area Under Precision-Recall Curve (AUPRC) (Boyd, Eng and Page 2013) can only be used to assess the misclassification detection performance of a single model rather than to compare directly between different models (Ashukha et al. 2020). To further reduce the bias of the results and ensure a fair comparison, extensive empirical evaluations are provided by employing the stratified nested cross-validation (CV) with the Tree-structured Parzen Estimator (TPE), which is one of the SoA hyperparameter optimisation algorithms.

4.2 Problem Statement

Reliability analysis for finger movement recognition relies on a framework that can explicitly measure the model reliability R , i.e., the quality of its uncertainty estimates. The challenges are manifold: it must be quantifiable and ideally located

in a fixed interval $[0, 1]$; it must be consistent for any classifier and uncertainty measure; the results must be comparable in a fair way regardless of the model accuracy. Inspired by studies on the evaluation of uncertainty quantification, the reliability of sEMG-based finger movement recognition could be evaluated by measuring the performance of the task of misclassification detection, which aims to detect wrong predictions with quantified uncertainty estimates as scores. An ideal reliable classifier enables the assignment of higher uncertainty measures when incorrect predictions are being made compared to correct predictions. In other words, the reliability R assesses the discrimination level of uncertainty quantification assigned to the wrong and correct predictions.

The misclassification detection can be considered as a binary classification problem where wrong predictions are positive samples and correct predictions refer to negative samples. The quantified uncertainty is taken as the score, and any samples with scores higher than a threshold will be assigned to positive samples, and negative ones otherwise. To avoid providing arbitrary results with a user-defined score threshold, the AUROC and AUPRC are commonly used as threshold-free evaluation summary metrics, which can overcome most of the challenges addressed above. However, these are incomparable since each model has its own accuracy in each test set, yielding different positive and negative samples regarding misclassification detection. More details of our proposed framework with a solution to address this challenge are presented in section 4.6.

4.3 Database

Our evaluations in this chapter were carried out on the Non-Invasive Adaptive Hand Prosthetics (NinaPro) Database 5 (*NinaPro DB5*), which was recorded with a double Myo setup in one session consisting of six repetitions of 52 hand movements (plus rest), which were divided into exercise sets A (finger movements), B

(hand and wrist movements), and C (other functional movements), performed by 10 healthy subjects (Atzori and Müller 2015). It should be noted that each repetition of all complete movements is sometimes referred to as *trial* (Wei et al. 2019) or *cycle* (Côté-Allard et al. 2019). Here, the term ‘cycle’ is employed to avoid confusion from the term ‘trial’ used in the hyperparameter optimisation process. Since we are particularly interested in sEMG-based finger movement recognition, only exercise A is used, which covers 12 finger movements that involve flexion and extension of five fingers plus thumb adduction and abduction. The illustration of these finger movements can be seen in Figure 25. To meet the real-time demands of controlling devices such as prostheses, i.e., the 300 ms constraint (Hudgins, Parker and Scott 1993), the raw sEMG data were segmented by applying a sliding window of 250 ms with a non-overlap length of 25 ms. This high overlap was used for data augmentation (Côté-Allard et al. 2019). Hence, each frame has a dimension of 16 electrode channels \times 50 sEMG sample points since the sampling frequency of NinaPro DB5 is 200 Hz. It should be noted that no additional signal preprocessing was performed here.

4.4 Models

To reduce bias, in our work, the *enhanced raw ConvNet* architecture, which was first proposed by (Côté-Allard et al. 2019), was used here to evaluate finger movement recognition performance in terms of both accuracy and reliability as a baseline method. It was modified to adapt for this task, which was to classify 12 finger movements by taking a frame of raw sEMG signals with a dimension of 16×50 . In essence, the CNN architecture was composed of two convolutional layers and two fully connected layers which have 2304 and 500 hidden units, respectively. The 3×5 kernels with a stride of 1 and valid padding were used in the convolutional layers. Note that valid padding is a type of padding in which

no padding is added to the input. By increasing the number of channels from 32 to 64, the convolutional layer is able to apply more complex filters and capture more abstract features of sEMG signals. Furthermore, recent techniques such as Batch Normalisation (BN) (Ioffe and Szegedy 2015), Parametric Rectified Linear Unit (PReLU) activation function (He et al. 2015b), and dropout were applied.

For a fair comparison, ECNN has the same network architecture as CNN except for the way of interpreting the model output and the loss functions used to train the network. It was designed by integrating CNN with Evidential Deep Learning (EDL) followed by the framework presented in Chapter 3. More details are shown in Figure 8.

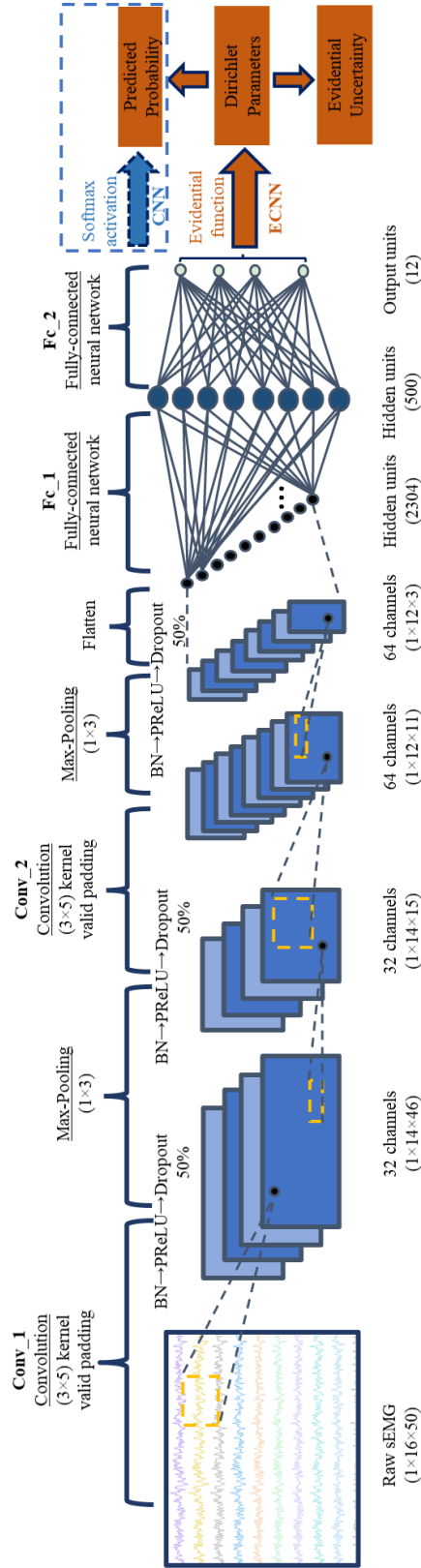


Figure 8: The detailed illustrations of the proposed ECNN and its conventional version (i.e., CNN). BN refers to Batch Normalisation. PReLU refers to Parametric Rectified Linear Unit. Conv_X (A×B) refers to the Xth convolution layer with (A×B) kernel valid padding. FC_X refers to the Xth Fully-Connected layer.

4.5 Illustration

This section aims to briefly illustrate the power of ECNN with its meaningful evidential uncertainty in classifying finger movements with raw sEMG. This was accomplished by conducting a comparison between ECNN-A and its conventional counterpart (CNN) through the use of equivalent main models, thereby guaranteeing a fair and straightforward evaluation. Briefly, the models were trained and tested only for the first subject from NinaPro DB5 to classify 12 finger movements with 16-channels raw sEMG signals, which was segmented using a 250 ms window with a 90% overlap. Therein, the models were trained by the 1st, 3rd, 4th and 6th cycles, whereas the 2nd cycle was used as a validation set for early stopping and the 5th cycle was used to test performance. To facilitate a more convenient comparison, we established a uniform training setting with a batch size of 256, a learning rate of 0.002, and the Adaptive Moment Estimation (ADAM) optimization method (Kingma and Ba 2015) for both ECNN-A and CNN models. Moreover, the cross-entropy loss was used for training the CNN, whereas the sum-of-squares loss as shown in Equation (14) was used for training the ECNN-A.

We first illustrate the power of the evidential uncertainty of ECNN-A by taking an example of classifying ‘thumb adduction’, which is easily confused during classification as ‘thumb flexion’ due to the similarity of movements. The top and bottom panels of Figure 9 show that CNN starts making wrong predictions during transient movements. This is consistent with the finding that the offline transient-state sEMG-based HGR accuracy is usually less than the steady-state one as the transient-state sEMG has more variance than the steady-state one over time (Englehart, Hudgins and Parker 2001; Jaramillo-Yáñez, Benalcázar and Mena-Maldonado 2020). The evidential uncertainty of ECNN reveals this clearly by presenting either high u_{vac} or u_{diss} during the transient phase, seen in the middle panel of Figure 9.

More importantly, it shows a clear understanding of the uncertainty sources

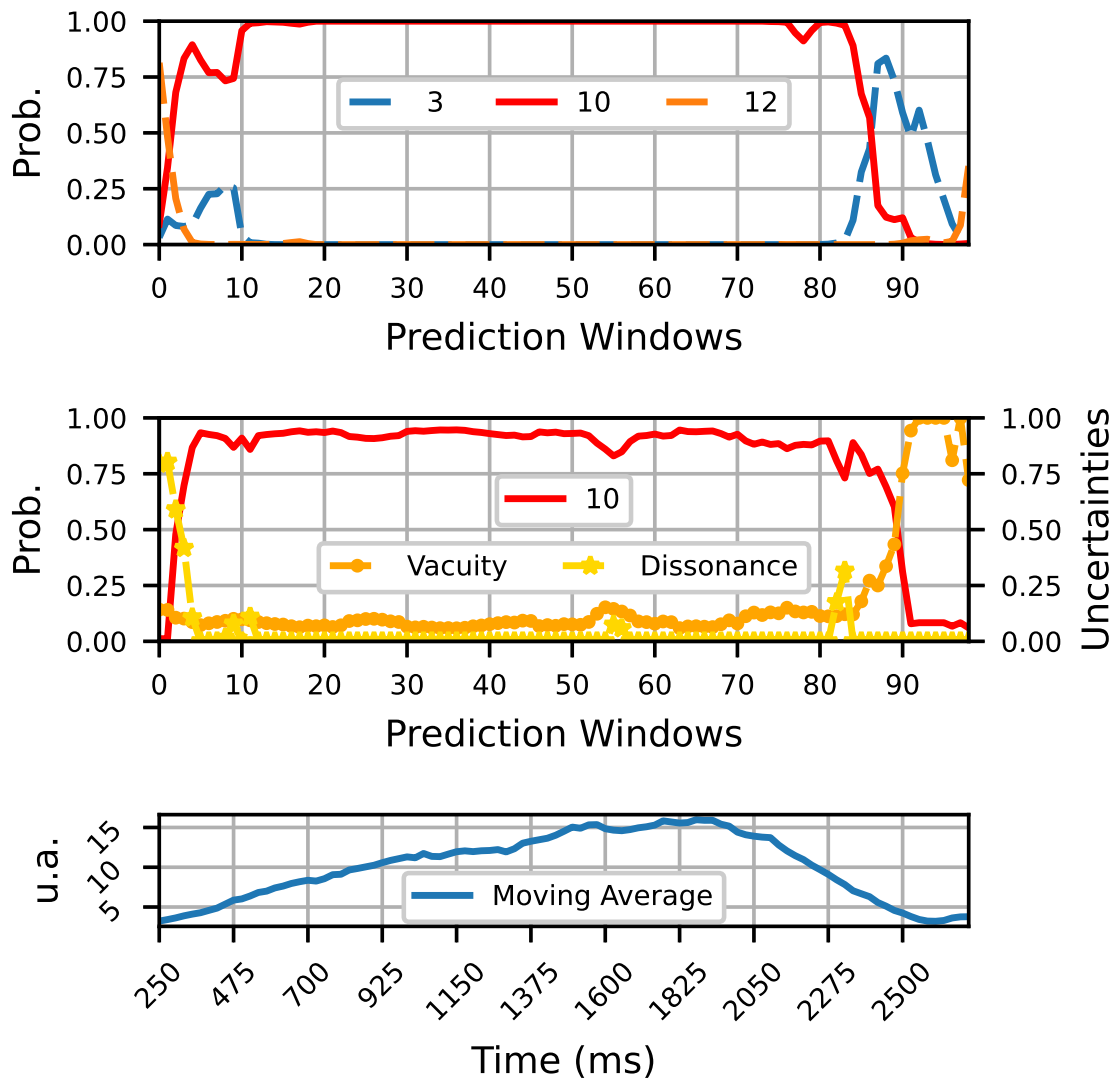


Figure 9: Sequential predictions of the ‘thumb adduction’ (class 10) on offline testing. Note that the sequential predictions of a wrong class are presented only if the predictive probability for at least one of them is above 0.5. Note that the ECNN refers to ECNN-A in this figure. **Top:** The predicted probabilities of the CNN; Class 3 and 12 refer to ‘middle flexion’ and ‘thumb flexion’. **Middle:** The predicted probabilities of the ECNN with its evidential uncertainty. **Bottom:** The sum of moving averages of 16-channel raw rectified sEMG signals with absolute values regarding the dynamic finger movement of ‘thumb adduction’. The u.a. denotes ‘unitless’ activation since sEMG recorded by Thalmic Myo armbands is claimed to be ‘unitless’ with an unknown conversion from mV.

in this example, which implies ECNN can provide better model interpretability. What CNN attempts to show is that the uncertainty at the beginning comes from conflicting evidence since its predicted probabilities for the 12th finger movement ‘thumb flexion’ are high at this stage. This is exactly what ECNN-A has revealed by giving high values of u_{diss} . Similarly, CNN shows ignorance at the end since it assigns high predicted probabilities for ‘middle flexion’, which seems unrelated to the ground truth ‘thumb adduction’. Again, this has been disclosed by ECNN-A via presenting high values of u_{vac} . From Figure 9, it can be seen that ECNN-A does not make overconfident predictions compared to CNN, especially when predictions may go wrong. Note that for ease of viewing, the focus is only on those classes with likely incorrect predictions, the sequential predictions of a wrong class will be presented in Figure 9 only when the probability for at least one of those wrong predictions is greater than 0.5. Based on this criterion, it can be observed that ECNN-A only displays the sequential predictions for the true class in Figure 9. This crucial evidence supports the assertion that ECNN-A avoids making overconfident predictions, at least for this example.

In summary, Figure 9 illustrates that ECNN-A has the potential to precisely quantify predictive uncertainties with understanding of the uncertainty sources. A natural question that arises is: how could we better leverage this for improving sEMG-based HGR performance? One straightforward solution is to allow a classifier to reject making a prediction when whichever dimension of uncertainty is considered as high. Assuming that high uncertainties are only generated when wrong predictions are made, making rejections under such conditions is then definitely a benefit to boost the HGR accuracy and make the accepted predictions more reliable. This is the intuition behind rejection-capable sEMG-based finger movement recognition. To briefly compare the classification performance of CNN and ECNN-A when allowing a model to reject making predictions by leveraging the uncertainty estimate, we first calculated $u_{nEntropy}$ for CNN and $max(u_{vac}, u_{diss})$

for ECNN regarding uncertainty estimates.

By setting a confidence threshold δ , where its range is set to $[0, 0.5]$, for discrimination between certain and uncertain predictions, the model is allowed not to make a prediction whenever its quantified uncertainty is larger than $(1 - \delta)$. When $\delta = 0$, it simply refers to the standard recognition where no rejections will be made. The upper limit of δ was set to 0.5 as a value of more than 0.5 is perceived to be too strict, which might lead to a situation in which no predictions are made. Inspired by studies of rejection-capable sEMG-based HGR, the three evaluation metrics used here are defined as follows: Rejection Rate (RR) is the percentage of predictions that are rejected (Nadeem, Zucker and Hanczar 2010; Scheme and Englehart 2015); True Acceptance Rate (TAR) refers to the rate at which a classifier correctly makes active predictions while True Rejection Rate (TRR) refers to the rate at which a classifier correctly makes inactive predictions. Note that False Acceptance Rate (FAR) and False Rejection Rate (FRR) were previously defined in (Scheme and Englehart 2015), and the sum of FAR/FRR and TAR/TRR equals to 1.

In terms of rejection-capable sEMG-based finger movement recognition, Figure 10 shows how ECNN-A outperforms CNN in this example. First, even though more predictions will be rejected as the confidence threshold δ increases, the blue lines show that the gradient of RR for ECNN-A is much smaller than CNN. When the threshold reaches 0.5, CNN almost stops making any predictions, but the RR of ECNN-A remains at about 10% only. This gives additional support to the proposed statement that CNN is considered overconfident. Second, the TAR of ECNN-A remains constant, whereas it drops for CNN as δ increases. Recall that TAR can be considered as accuracy in finger movement recognition but under the condition that the model is allowed not to make uncertain predictions. The standard recognition accuracy of ECNN-A is also higher than CNN, as shown in pink points when $\delta = 0$.

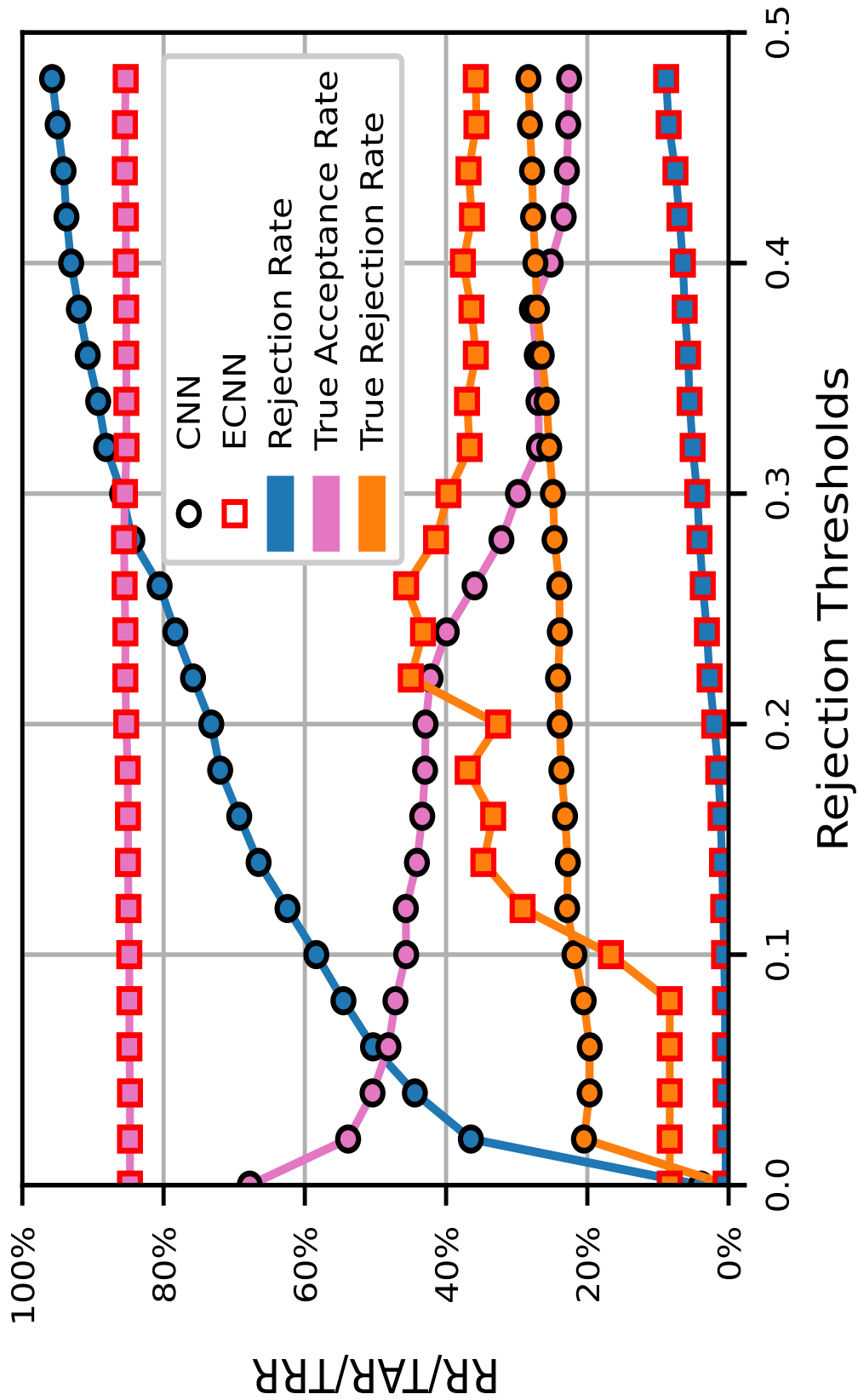


Figure 10: Three comparison results of classifying 12 finger movements between CNN and ECNN when rejecting uncertain predictions. Note that the ECNN refers to ECNN-A in this figure.

Finally, it shows that ECNN-A is making more valid rejections generally than CNN, supported by the TRR shown in orange. It is observed that ECNN-A has a lower TRR when the δ varies from 0 to 0.1, which may be caused by the extremely low RR of ECNN-A, that is, very few predictions are rejected when δ is small.

Although ECNN-A has shown its superiority in this example, it must be understood that one example alone may not prove that ECNN-A is more reliable than CNN. Therefore, the illustration here can only be considered as a supplement for the readers to better understand the special properties of ECNN-A with evidential uncertainty. This small example also indicates how to investigate the rejection-capable sEMG-based finger movement recognition performance with uncertainty measures in a traditional way. The proposed proper reliability analysis for both models will be presented in detail later.

4.6 Experimental Setup

All experiments were implemented in PyTorch *v.1.1.0* and Python 3.7.3. The experimental sequences were constructed by data loading, data segmentation, model training, and model testing. A standard CV procedure may cause biased results when assessing classification models (Varma and Simon 2006; Krstajic et al. 2014). To reduce bias and to better compare finger movement recognition performance between CNN and ECNN, a stratified nested CV procedure (Stone 1974; Krstajic et al. 2014) was employed in this work, where an inner CV loop was used to determine the best hyperparameters for model training, while an outer CV was then applied to test and compare the results. Stratification allows each fold divided from the data to have similar proportions of samples with the same label. This could be done by simply splitting the data via the repetition number here. The detailed algorithm for model training with stratified nested CV can be seen in Algorithm 1.

Algorithm 1 Model Training with Stratified Nested CV

Input: $\mathbf{D} = \{\mathbf{X}_i, \mathbf{Y}_i\}_{i=1}^N$, dataset includes segmented raw sEMG signals with labels, which has been divided by the repetition number from 1 to N . Define loss function J .

Output: Model parameters $\theta = \{\theta_1, \dots, \theta_N\}$ after training

```

1: for Each Repetition  $i$  do {Outer Loop CV}
2:   Load testing set  $\mathbf{D}^{(test)} = \{\mathbf{X}_i^{(test)}, \mathbf{Y}_i^{(test)}\}$ 
3:   (Hyperparameter Optimisation)
4:   for Each trial of hyperparameter study  $k$  do
5:     Define the objective function of hyperparameter study with proposed hyperparameter search space
6:     Initialise a list  $\mathbf{O}$  for collecting the objective values
7:     Initialise a list  $\mathbf{L}$  for collecting the validation losses from the inner loop CV {Inner Loop CV}
8:     for Each repetition  $j$  ( $j$  not  $i$ ) do
9:       Load validation set  $\mathbf{D}^{(val)} = \{\mathbf{X}_j^{(val)}, \mathbf{Y}_j^{(val)}\}$ 
10:      Let the remaining dataset be the training set  $D^{(train)}$ 
11:      Initialise  $\theta_{ij}$  with random values
12:       $best\_val = inf$ 
13:      for Each epoch do
14:        Update the  $\theta_{ij}$ 
15:        if  $J(\mathbf{X}^{(val)}, \mathbf{y}^{(val)}) < best\_val$  then
16:           $best\_val = J(\mathbf{X}^{(val)}, \mathbf{y}^{(val)})$ 
17:           $counter = 0$ 
18:        else
19:           $counter += 1$ 
20:        end if
21:        Stop training when  $counter$  reaches to 10
22:      end for
23:      if the pruning is activated then
24:        Break the inner loop and move to the next hyperparameter study trial
25:      else
26:        Add  $best\_val$  to the list  $\mathbf{L}$ 
27:      end if
28:    end for
29:    Add the objective value  $mean(\mathbf{L})$  to the list  $\mathbf{O}$ 
30:  end for
31:  Load retraining dataset by combining both  $\mathbf{D}^{(val)}$  and  $\mathbf{D}^{(train)}$ ,
  i.e.,  $\mathbf{D}^{(retrain)} = \{\mathbf{X}^{(retrain)}, \mathbf{y}^{(retrain)}\}$ 
32:  Initialise the model parameters  $\theta_i$  with random values
33:  Apply the optimal hyperparameter set which yields  $min(\mathbf{O})$ 
34:  for Each epoch do
35:    Update the  $\theta_i$ 
36:    Stop training when  $J(\mathbf{X}^{(retrain)}, \mathbf{y}^{(retrain)})$  reaches to  $min(\mathbf{O})$ 
37:  end for
38:  Save model parameters  $\theta_i$  to  $\theta$ 
39: end for
40: return  $\theta$ 

```

Since each subject performed six repetitions of all gestures in the NinaPro DB5, the split ratio of training, validation, and testing datasets was set to 4 : 1 : 1 regarding cycle number to maximise the data used for training. This data split method could also prevent data leakage between training and testing. Recall that the raw sEMG signal was segmented by a sliding window and the overlap between every two consecutive frames was as high as 90%. Therefore, randomly splitting the sample set may cause such a leakage scenario in which a sample falls into the training set while its adjacent segments could be found in the testing set. Furthermore, early stopping was employed to avoid overfitting by setting the patience term at 10. Specifically, to ensure the training process did not continue unnecessarily, we stopped the training when no improvement was detected in the validation set after waiting for 10 epochs, or when the training epoch reached 1000.

Unlike conventional Hyperparameter Optimisation (HPO) algorithms such as Grid or Random Search, we applied one of the SoA HPO algorithms, the TPE (Bergstra et al. 2011; Bergstra, Yamins and Cox 2013), to reduce the computation burden. Being an approach based on sequential model-based global optimisation algorithms (Hutter, Hoos and Leyton-Brown 2011; Bergstra et al. 2011), the TPE organises hyperparameters into a tree-like space so that the available values of a specific hyperparameter will be determined based on the previous search results. With the help of Optuna (Akiba et al. 2019), which is a powerful hyperparameter optimisation framework, the unpromising *trials* will be terminated at an early stage where each trial refers to each evaluation of an objective function. Such a strategy is also referred to as *pruning*, and the ‘MedianPruner’ constructed by the Median Stopping Rule (Golovin et al. 2017) was used here. Specifically, the objective value is then the mean of the validation losses collected from the inner CV loops. Moreover, the number of study trials was set to 25 and the pruning was enabled after 5 trials were completed

in each process of HPO. The source code for this study is available on GitHub (<https://github.com/YuzhouLin/ECNN-RAnal>), and the determined optimal hyperparameters of each model on each test trial of CV for each individual can be found here as well.

Table 1: Hyperparameter Search Space

Hyperparameters	CNN	ECNN-A	ECNN-B	ECNN-C
batch size		{128, 256}		
learning rate		[1e-3, 1e-2]		
optimizer		{"ADAM", "RMSprop", "SGD"}		
evidence func	×	{"ReLU", "SoftPlus", "Exp"}		
s	×	×	[10, 60]	×
λ	×	×	×	[0.1, 1.0]

The hyperparameter search space is listed in Table 1. The common hyperparameters used for training both CNN and ECNN include batch size, learning rate, and optimiser method. To better explore the potential of ECNN, we investigated different functions to generate the evidence vector (called ‘evidence func’ in Table 1) and train the model. Instead of employing *ReLU* as the last activation function for ECNN to turn the model outputs into the nonnegative evidence vector for the predicted Dirichlet distribution, other functions such as *SoftPlus* and the exponential function (*Exp*) can be investigated. Note that any value larger than 3 would be limited to 3 when using the exponential function for training convergence. Recall that ECNN-B and ECNN-C are trained by incorporating a Kullback-Leibler (KL) divergence term into the sum-of-squares loss function. Based on the training strategies employed, a specific hyperparameter for ECNN-B and ECNN-C is the annealing step s and λ , respectively. Eventually, the determined optimal hyperparameters of models employed for reliability analysis of finger movement recognition can be found in Appendix B.

4.7 Performance Evaluation

4.7.1 Evaluation of Accuracy

First, we use the recall to evaluate the general efficacy of sEMG-based finger movement recognition. As a multiclass classification problem, recall can be calculated by taking the macroaverage and microaverage. The macroaverage recall is calculated as:

$$r_M = \frac{1}{K} \sum_{j=1}^K \frac{tp_j}{tp_j + fn_j}, \quad (19)$$

where r_M is the macroaverage recall; tp and fn represent the number of true positives and false negatives; K is the number of finger movements and j refers to a specific one. It was employed here to measure the average per-class accuracy of such recognition because each finger movement is considered equally important, whereas the microaverage one favours bigger classes (Sokolova and Lapalme 2009). It would be further averaged over subjects for overall comparison. Second, to further investigate the accuracy of rejection-capable sEMG-based finger movement recognition, and for the sake of consistency with its related studies, the evaluation metric of Accuracy-Rejection Curve (ARC) (Nadeem, Zucker and Hanczar 2010; Scheme and Englehart 2015) was used here to compare the performance of CNN and ECNN variants in terms of their rejection rates. By varying the rejection threshold δ from zero to one, different pairs of RR and the corresponding accuracy (i.e., TAR) could be achieved when testing a trained classifier. For the overall comparison, we calculated the mean ARC for each model using the 20 bins of RR under the CV scheme.

4.7.2 Evaluation of Reliability

As indicated in section 4.2, the reliability of the sEMG-based finger movement recognition could be evaluated by measuring the performance of the misclassification detection. The AUROC and AUPRC can then be used to calculate the model reliability and are noted as R_{AUROC} and R_{AUPRC} , which can be simply computed using the trapezoidal rule and Average Precision (AP) shown in (20), respectively. Consider a testing data set $\mathbf{D}^{(test)}$ with n samples and the number of positive (incorrect predictions) and negative samples (correct predictions) is represented by n_{pos} and n_{neg} , respectively,

$$AP = \frac{1}{n_{pos}} \sum_{i=1}^{n_{pos}} p(i), \quad (20)$$

where n samples will be sorted from high to low based on uncertainty estimates and i is the rank in the sequence of sorted positive samples; $p(i)$ is the precision at cut-off i . It has been proven to be one of the most robust estimators for summarising information in Precision-Recall Curve (PRC) (Boyd, Eng and Page 2013).

Since each model has a specific class skew π in the detection of misclassification, defined as n_{pos}/n , it is inappropriate to use R_{AUROC} and R_{AUPRC} for direct comparison between models. We recommend measuring the model reliability by R_{nAUPRC} for a robust and fair comparison, which is a normalised AUPRC. In this paper, we will present the results of R_{AUROC} and R_{AUPRC} for all models as a reference only and those of R_{nAUPRC} for performance comparison. Boyd et al. first proved that there is a region of PRC that is not achievable, and the area of such an unachievable region depends on π . The nAUPRC was therefore proposed to account for this using normalisation. As such,

$$R_{nAUPRC} = \frac{AP - AP_{min}}{AP_{max} - AP_{min}}, \quad (21)$$

where $AP_{max} = 1$, that is, the theoretical maximum AUPRC; AP_{min} is the theoretical minimum AUPRC proved by (Boyd et al. 2012), see below:

$$AP_{min} = \frac{1}{n_{pos}} \sum_{i=1}^{n_{pos}} \frac{i}{n_{neg} + i}, \quad (22)$$

where n_{pos} and n_{neg} represent the number of positive and negative samples; i is the rank in the sequence of sorted positive samples.

4.7.3 Evaluation under Cross-Validation

There are two incompatible ways to compute the proposed evaluation metrics under nested CV. It can be calculated by either taking the mean of the results from each fold in the outer loop CV or aggregating the data from all folds into one first and then followed by the equations. Since merging assumes that the models are calibrated (Forman and Scholz 2010), which is not the case here, all evaluation metrics will be computed using the former approach here.

4.8 Results

In all experiments, unless otherwise stated, the performance of CNN is taken as the baseline and compared with ECNN variants using statistical analysis with the Wilcoxon signed rank test, where the null hypothesis assumes that there is no difference in the evaluation results between the two models and will be rejected when the p-value < 0.05 . The difference in performance among ECNN variants will also be investigated.

4.8.1 Accuracy Analysis

Here we first verified the accuracy of CNN and three ECNN variants.

Without rejection

It can be seen that ECNN-A and ECNN-C outperformed CNN overall in terms of classification accuracy on the NinaPro DB5 in Table 2. The average improvements, which were statistically significant, reached 1.72% and 1.46% respectively. It should be noted that the difference of accuracy between ECNN-A and ECNN-C was not statistically significant, and CNN significantly outperformed ECNN-B but with a difference of only 2.17% on accuracy. As such, one could notice that the rank of model accuracy was $ECNN-A \approx ECNN-C > CNN > ECNN-B$. Details of the model performance from the perspective of nested CV are shown in Figure 11 and the same finding can be observed on each fold in the outer loop CV from it. It is interesting to note that all models achieved the lowest accuracy in the 1st fold, indicating that there is significant variability between the first trial of sEMG and others. This may be because subjects need time to accommodate the Myo band to perform hand gestures.

Table 2: Macroaverage Recall of the ConvNets with Comparisons

Models	ECNN-A	ECNN-C	CNN	ECNN-B
M \pm STD (%)	76.34 \pm 21.1	76.08 \pm 20.9	74.62 \pm 21.7	72.45 \pm 22.3
H0 (p) *	-	1 (7e-02)	0 (2e-06)	0 (6e-22)
H0 (p)	0 (2e-06)	0 (7e-05)	-	0 (2e-04)

* The Wilcoxon signed rank test was employed to compare the ECNN-A, which yields the highest mean macro average recall, with other models. The same test was used to compare the CNN with ECNN variants, where the statistical results are presented below this line.

- The M and STD refer to the mean and standard deviation of macroaverage recall under nested cross-validation over 10 subjects.

Furthermore, the average confusion matrices for CNN and three ECNN variants are presented in Figure 12, where each annotated score represents the per-class normalised accuracy averaged over six outer CV trials across 10 subjects. It can be seen that all models perform similarly. For example, they all performed well in the classes ‘2 (Middle flexion)’, ‘3 (Middle extension)’, ‘7 (Little finger

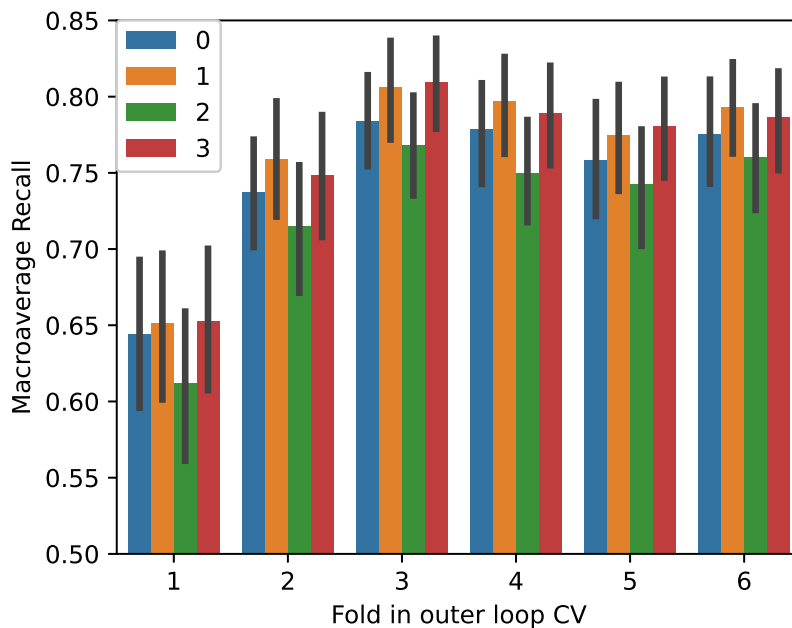


Figure 11: Accuracy comparisons of the CNN and ECNN variants with nested cross validation. 0 is CNN and 1 – 3 refer to ECNN-A, ECNN-B and ECNN-C.

extension)', '9 (Thumb adduction)' and '11 (Thumb flexion)', while the pair (8, 10) is found to be more closely related than the other classes. Note that the class 8 ('Thumb abduction') and the class 10 ('Thumb extension') are commonly confused with each other. Regarding the per-class performance comparison of models for finger movement recognition, it can be observed that ECNN-A and ECNN-C performed better than CNN and ECNN-B in all classes except 'Ring flexion' (class 4) and 'Thumb extension', where ECNN-C achieved a slightly lower accuracy than CNN in these two classes, with the differences of 0.05% and 0.26% only. Furthermore, CNN outperformed ECNN-B in most classes except for 'Middle extension', 'Ring extension' (class 5), and 'Thumb adduction'.

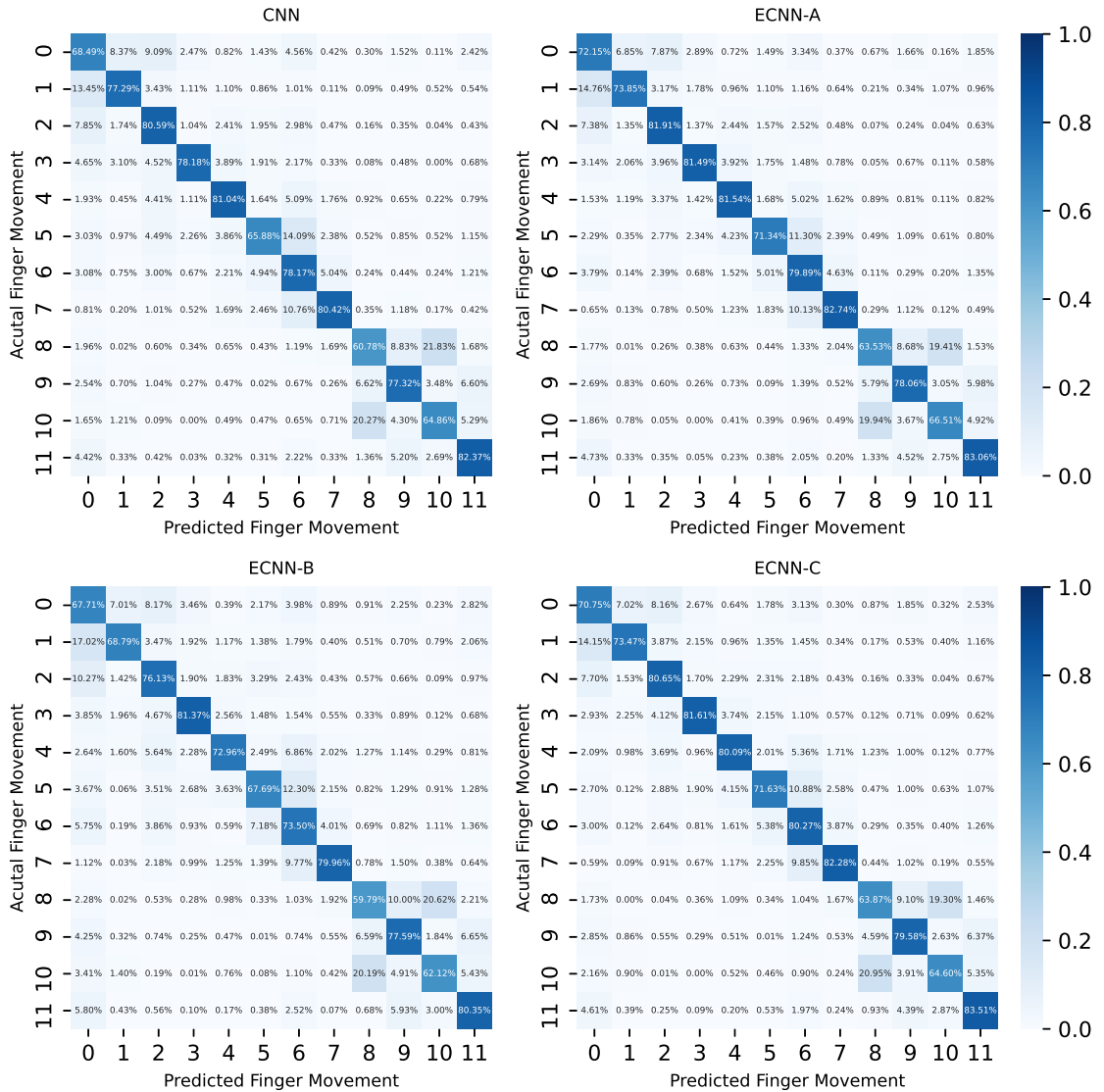


Figure 12: Confusion matrices averaged over all subjects with nested cross validation of the CNN and ECNN variants. Note that 0, 2, 4, 6, 10 are the flexion movements of finger Index, Middle, Ring, Little finger and Thumb; The extension movements of each finger are then represented as class 1, 3, 5, 7, 11. Furthermore, 8, 9 refer to Thumb adduction and abduction.

With rejection

As shown in Figure 13, the recognition accuracy comparison of rejection schemes in the form of ARC reveals the trade-off relationship between the proportion of rejections and the resulting accuracy of the active predictions. It could be clearly seen that ECNN-A was not substantially greater than ECNN-C and both outperformed CNN and ECNN-B in terms of recognition accuracy under the rejection condition, where the latter two also had approximately the same performance. With a specific focus on the regions where the models had low RRs (i.e., $0 < RR \leq 15\%$), which may be a reasonable target range in practical scenarios, all ECNN variants obtained higher accuracy than CNN.

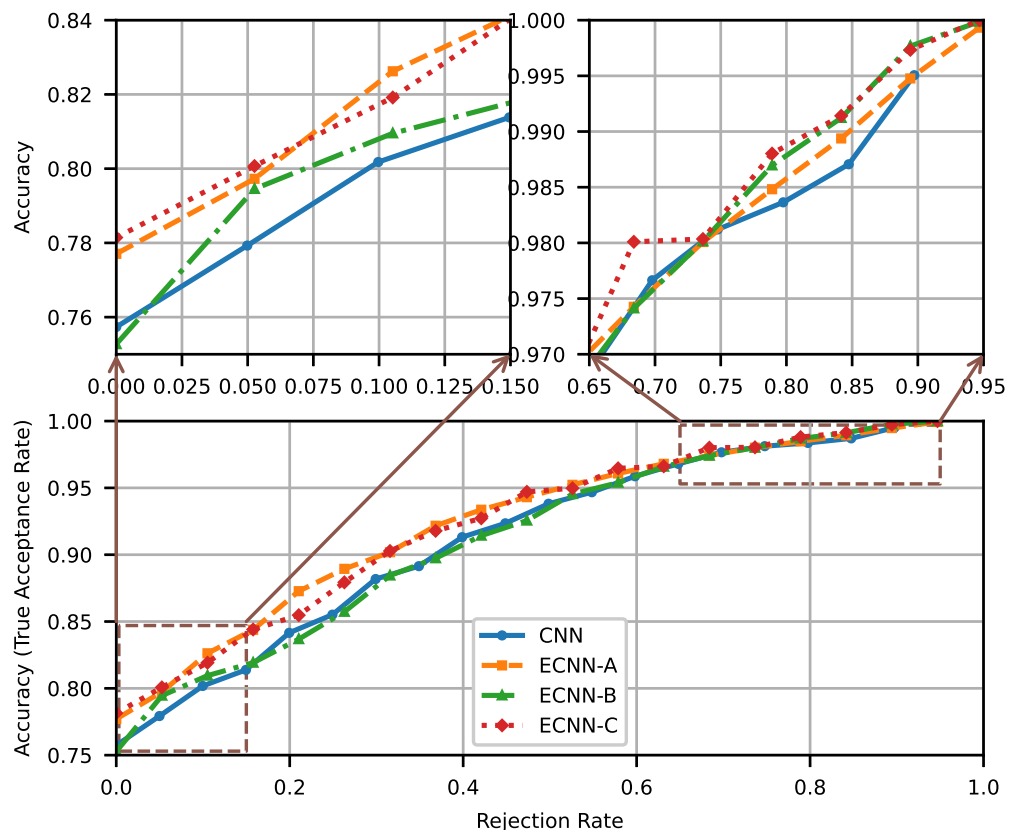


Figure 13: The mean ARC plots of all models under CV scheme when considering the ‘overall’ uncertainty estimate.

4.8.2 Reliability Analysis

Here, we investigate the reliability analysis of CNN and three ECNN variants regarding different uncertainty estimates. In addition to considering evidential uncertainties including u_{vac} and u_{diss} specifically for ECNN variants, the utilisation of entropy and negative maximum probability serves as two common uncertainty measures for all models (Scheme and Englehart 2015).

The entropy H for the sample i can be simply defined as:

$$H_i = - \sum p_{ij} \ln p_{ij}, \quad (23)$$

where p_{ij} represents the predicted probability assigned by the model that sample i belongs to class j . Since the maximum probability across classes can be interpreted as the confidence level, it could be used as an uncertainty score by taking its negative value, i.e., $-max(\mathbf{p})$. However, the range of entropy and negative maximum probability is $[0, \ln(1/K)]$, and $[-1, 0]$ respectively, where K represents the number of classes. For consistency, they will be normalised to a range from 0 to 1 and noted as $u_{nEntropy}$ (normalised Entropy) and u_{nnmp} (normalised negative maximum probability). Furthermore, from the perspective of practical use, the overall uncertainty was noted as ‘overall’ in Table 3 and calculated by $max([u_{nEntropy}, u_{nnmp}])$ for CNN and $max([u_{nEntropy}, u_{nnmp}, u_{vac}, u_{diss}])$ for ECNN variants. Recall that the reliability analysis directly measures the quality of uncertainty estimates and only R_{nAUPRC} can be used for performance comparison between models.

Table 3: Reliability Comparison of the ConvNets by Evaluating the Misclassification Detection regarding Uncertainty Estimates

Scores	Models	R_{AUROC}		R_{AUPRC}		R_{nAUPRC}	
		M(%) \pm STD(%)	H0 (p)	M(%) \pm STD(%)	H0 (p)	M(%) \pm STD(%)	H0 (p)
$u_{nEntropy}$	CNN	82.06 \pm 3.72	-	54.71 \pm 8.57	-	47.64 \pm 8.05	-
	ECNN-A	83.17 \pm 4.72	0 (4.0e-03)	55.69 \pm 8.82	1 (5.9e-01)	49.32 \pm 8.82	1 (2.7e-01)
	ECNN-B	84.41 \pm 4.76	0 (1.9e-06)	63.37 \pm 8.65	0 (1.9e-08)	56.85 \pm 8.94	0 (2.6e-08)
	ECNN-C	83.65 \pm 4.59	0 (3.2e-04)	57.64 \pm 7.98	0 (1.1e-02)	51.39 \pm 7.98	0 (3.0e-03)
u_{nmp}	CNN	83.18 \pm 3.68	-	57.09 \pm 7.05	-	50.33 \pm 6.37	-
	ECNN-A	83.60 \pm 4.46	1 (2.0e-01)	56.62 \pm 7.23	1 (2.3e-01)	50.35 \pm 6.91	1 (5.6e-01)
	ECNN-B	84.70 \pm 4.61	0 (2.4e-04)	63.80 \pm 8.28	0 (1.4e-07)	57.35 \pm 8.53	0 (1.5e-07)
	ECNN-C	83.92 \pm 4.62	0 (1.1e-02)	58.24 \pm 6.73	1 (3.1e-01)	52.02 \pm 6.65	1 (1.2e-01)
u_{vac} *	ECNN-A	71.88 \pm 7.51	-	44.96 \pm 11.48	-	37.03 \pm 11.70	-
	ECNN-B	83.02 \pm 5.20	0 (1.6e-11)	62.04 \pm 9.36	0 (3.5e-11)	55.32 \pm 9.68	0 (3.8e-11)
	ECNN-C	76.64 \pm 8.42	0 (6.5e-07)	51.22 \pm 9.77	0 (3.0e-08)	44.09 \pm 9.80	0 (2.4e-08)
u_{diss} *	ECNN-A	77.32 \pm 5.13	-	44.96 \pm 7.61	-	36.98 \pm 6.82	-
	ECNN-B	70.50 \pm 8.01	0 (3.9e-10)	42.06 \pm 8.85	0 (2.7e-03)	31.61 \pm 8.41	0 (1.0e-05)
	ECNN-C	74.75 \pm 5.41	0 (8.7e-06)	42.58 \pm 6.27	0 (5.3e-04)	33.99 \pm 5.32	0 (1.5e-04)
overall	CNN	82.76 \pm 3.74	-	55.96 \pm 8.03	-	49.06 \pm 7.50	-
	ECNN-A	83.47 \pm 4.60	0 (2.9e-02)	55.07 \pm 8.18	1 (1.7e-01)	48.56 \pm 8.11	1 (3.5e-01)
	ECNN-B	84.42 \pm 4.77	0 (1.8e-04)	63.37 \pm 8.65	0 (1.2e-07)	56.86 \pm 8.94	0 (2.9e-07)
	ECNN-C	83.85 \pm 4.46	0 (2.6e-03)	57.82 \pm 7.75	1 (8.2e-02)	51.60 \pm 7.70	0 (3.4e-02)

* The Wilcoxon signed rank test was employed to compare the ECNN-A with other ECNN variants.

- The M and STD refer to the mean and standard deviation of reliability measures under nested cross-validation over 10 subjects.

- Note that if the statistical results are conflicting, the results of R_{nAUPRC} shall be the standard ones.

From Table 3, our first finding regarding the quality of uncertainty estimates was that all models with the uncertainty estimate u_{nnmp} achieved an overall highest R measured by either R_{AUROC} , R_{AUPRC} , or R_{nAUPRC} compared to other types of uncertainty estimate. Moreover, ECNN variants with the uncertainty estimate of either u_{vac} or u_{diss} alone generally obtained poor results of R . Our second finding regarding the R comparison between CNN and ECNN variants was that ECNN-B significantly outperformed CNN under any condition, where the highest improvement of reliability R_{nAUPRC} of 19.33% was achieved with the uncertainty estimate $u_{nEntropy}$ and 15.90% for the ‘overall’ uncertainty estimate. However, the difference in R_{nAUPRC} between CNN and ECNN-A was not significant under any condition, while that between CNN and ECNN-C was not with u_{nnmp} alone. Regarding the comparison of ECNN variants, ECNN-B achieved the highest R when using vacuity as the uncertainty estimate. Despite ECNN-A performed best when using dissonance as the score of misclassification detection, the results of R_{nAUPRC} for all ECNN variants were generally quite low (no more than 36.98%). Eventually, the observed order of R_{nAUPRC} obtained with the uncertainty estimate of ‘overall’ was ECNN-B > ECNN-C > ECNN-A \approx CNN.

4.9 Discussion

The current study had a particular focus on improving model efficiency and robustness, but not directly investigating model reliability. To fill this gap, we defined the model reliability R as the quality of its uncertainty estimate and proposed an offline framework to quantify it. We focused our examination on the model reliability, and one implication of the results is that ECNN has great potential for complex and versatile finger movement recognition. Specifically, ECNN-C outperformed CNN with $p < 0.05$ in both accuracy and reliability with a difference of 1.46% in r_M (Table 2), and 2.54% in R_{nAUPRC} with the ‘overall’

uncertainty (Table 3), respectively. This suggests that training ECNN with a constant effect of KL should be applied when both model efficiency and reliability are weighted equally. Additionally, the loss function excluding the KL term is suggested for training the ECNN if the efficiency of the model matters more than the reliability. This is supported by the finding that ECNN-A achieved the best r_M of 76.34%, which was 1.72% higher than CNN with $p < 0.001$ (Table 2) - but no significant differences of R_{nAUPRC} were found between them (Table 3).

Note that ECNN-A has shown its efficiency by presenting the SoA performance on NinaPro DB5 (Exercise A) since the best accuracy reported in the literature was 76.02%, achieved by taking an input of 300 ms sEMG signals to an ensemble classifier of three CNNs (Shen et al. 2019). On the contrary, ECNN is recommended to be trained by taking the annealing effect of KL term when there is a serious concern about model reliability, e.g., controlling a prosthetic limb for daily tasks to meet the needs of transradial amputee users. Our findings indicate that ECNN-B was determined to be the most reliable one by showing improvements ranging from 14.25% to 19.33% in R_{nAUPRC} with different uncertainty measures (Table 3), compared to CNN. Although it was found to be less accurate than CNN where the difference in r_M was approximately 2% (Table 2), its accuracy under the rejection scheme was approximately equal to CNN in general and even better than CNN when RR is in a low range of 0% to 15% (Figure 13).

Defining the comparable model reliability has implications for understanding how much an sEMG-based hand gesture classifier knows about its predictions, thereby providing us with general guidelines for designing such a reliable model which has the potential to improve its efficiency by rejecting wrong predictions with the aid of its uncertainty estimate. The proposed framework of reliability analysis measures R by evaluating the performance of misclassification detection using the score of uncertainty estimate. Therefore, a model with a higher R could generate more discriminate uncertainty estimates, i.e., lower uncertainty estimates

are assigned to correct predictions and vice versa. This implies that the value of R indicates how easily an optimal rejection threshold can be found for rejection-capable sEMG-based HGR. By measuring it, one can easily check the reliability of a model without the need to test its performance when allowing rejection by measuring several evaluation metrics such as RR, TAR, and TRR across a range of rejection thresholds.

Furthermore, we highly recommend using nAUPRC to measure R , although AUROC and AUPRC are commonly used to test the performance of a misclassification detection task. The following order could be observed in each reliability analysis of a model with an uncertainty estimate: $R_{AUROC} > R_{AUPRC} > R_{nAUPRC}$. This finding is consistent with other research showing that Receiver Operating Characteristics (ROCs) usually make innocent impressions, whereas PRCs reveal the bitter truth, especially in unbalanced data sets (Saito and Rehmsmeier 2015). We argue that the general low value of R_{nAUPRC} may just exactly represent the situation in reality, since the averaging of nAUPRC under CV can further reduce the effect of the skew (Boyd et al. 2012).

There are a few limitations that are important to note. First, one cannot investigate the R of a model when it is tested with a classification accuracy of 100% or 0% because there are no positive or negative samples for misclassification detection in this case. We suggest setting R to 0 since such unusual results imply that the model needs to be further investigated and cannot be easily trusted. Second, even though we have demonstrated the potential of ECNN, the implications of its understandable evidential uncertainty remain to be explored. Hypothetically, understanding the source of uncertainty is helpful to improve the robustness of the model by making valid rejections. A potential research direction would then be to investigate the relationship between the proposed reliability analysis and current studies on model robustness. Third, measuring the performance of misclassification detection with nAUPRC may not be the only way to investigate R .

For example, it could be investigated by computing the area under ARC or measuring the performance of out-of-domain data detection (e.g., unseen gestures or adversarial samples). We encourage researchers to address the problem of sEMG-based HGR from the perspective of model reliability together with model efficacy and robustness.

4.10 Summary

This chapter has raised concerns about the reliability of the sEMG-based HGR models. The great potential of our proposed uncertainty-aware model, i.e. ECNN, has been demonstrated to classify 12 individual finger movements. This was done using the reliability analysis, which is based on the proper definition of the model reliability R we provided. In short, the R of a model is defined as the quality of its inferred uncertainty measures. The experimental results on a benchmark dataset, *NinaPro DB5* (Exercise A), show that ECNN is more reliable than CNN in general, where the highest reliability improvement of 19.33% is observed. Furthermore, it has been illustrated that the best mean accuracy achieved by ECNN over 10 subjects is 76.34%, which is slightly higher than the SoA performance. This work demonstrates the potential of ECNN and recommends using the proposed reliability analysis as a supplementary measure to evaluate sEMG-based HGR.

Chapter 5

Improve the Long-Term Robustness of Hand Grasp Recognition

5.1 Introduction

Hand Gesture Recognition (HGR) with Surface Electromyography (sEMG), which reveals neuromuscular activities by collecting electrical signals from muscles via non-invasive electrodes, has been widely acknowledged as a natural way to express intuitive intention, thus providing a control command for Human-Machine Interaction (HMI) (Hakonen, Piitulainen and Visala 2015; Jaramillo-Yáñez, Bernalcázar and Mena-Maldonado 2020). As such, designing an accurate, robust, and reliable hand gesture classifier can help build a solid bridge between computers and humans. This is extremely valuable for transradial amputees in controlling prosthetic limbs, allowing them to perform Activities of Daily Living (ADL) through hand movements, including various hand grasps. As an independent module in

HMI, sEMG-based HGR has often been studied offline with publicly available datasets (Atzori and Müller 2015; Cognolato et al. 2020; Côté-Allard et al. 2021).

Recently, deep learning approaches have been extensively investigated to improve the performance of sEMG-based HGR, and state-of-the-art (SoA) results have been reported on a series of benchmark data sets achieved by different deep learning models, particularly Convolutional Neural Networks (CNNs) (Wei et al. 2019; Côté-Allard et al. 2019). However, the challenge remains to obtain the long-term robustness of the sEMG-based HGR. As many have reported, the recognition accuracy of a well-trained sEMG-based hand gesture classifier may vary or even decrease significantly over time (He et al. 2015a; Zia ur Rehman et al. 2018a; Côté-Allard et al. 2021). Various possible factors responsible for the poor long-term robustness of HGR have been studied and could be summarised as muscle fatigue, skin conductivity, limb position, electrode displacement, and signal variation over days (Hargrove, Englehart and Hudgins 2006; Kaufmann, Englehart and Platzner 2010; Scheme et al. 2010; Scheme and Englehart 2011; Young, Hargrove and Kuiken 2012). To address these issues, we believe that rejecting uncertain predictions (Robertson, Englehart and Scheme 2019; Wu et al. 2021; Bao et al. 2022) is the most straightforward approach compared to others such as multimodal approaches (Krasoulis et al. 2017; Wei et al. 2019; Cognolato et al. 2022), adaptive learning (Tommasi et al. 2013; Vidovic et al. 2016; Zhu et al. 2017; Betthausen et al. 2018; Côté-Allard et al. 2020b), and alternative training protocols (Beaulieu et al. 2017; Yang et al. 2017a; Zanghieri et al. 2020a; Côté-Allard et al. 2020b).

Our previous work (Lin et al. 2022) has investigated the reliability (i.e., the quality of uncertainty measures) of the uncertainty-aware models that were proposed for the recognition of finger movement with raw sEMG. Furthermore, we have shown that a classifier can be considered more reliable if it knows what it does not know. The promising results encouraged us to pursue further research on the long-term robustness of hand grasp recognition with the same design of

uncertainty-aware models. In this study, we focus on the very challenging Non-Invasive Adaptive Hand Prosthetics (NinaPro) Database 6, as it (i) was first released for the repeatability analysis of hand grasp recognition with sEMG collected in data acquisitions across days; (ii) was found to require more investigations compared to other benchmark data sets; (iii) was ideal for validating the long-term robustness of the proposed uncertainty-aware models on sEMG-based hand grasp recognition. We first design end-to-end 2D CNN and 3D CNN as baseline models for the specific task of classifying hand grasps. Then uncertainty-aware models, that is, 2D Evidential Convolutional Neural Network (ECNN) and 3D ECNN, are proposed by integrating them with Evidential Deep Learning (EDL) respectively.

Despite the fact that the potential of ECNN can be investigated by reliability analysis without optimal rejection threshold determination, one may be eager to know how the performance of HGR can be improved under rejection schemes in practical use. Therefore, the primary objective of this study is to present a practical way to determine a rejection threshold using only validation sets and to provide a comprehensive analysis of the long-term performance of rejection-capable hand grasp recognition with raw sEMG. Furthermore, ECNN could generate multidimensional uncertainties such as vacuity and dissonance as a result of the nature of EDL (Lin et al. 2022). A secondary objective of this study is to investigate its potential to improve the rejection-capable performance with multidimensional uncertainties by comparing it with a single uncertainty.

5.2 Related Work

The NinaPro DB6 (Palermo et al. 2017) is an undervalued benchmark data set to test the repeatability of hand grasp recognition, where repeatability was defined as the variation in repeated measurements made on consecutive days by the same

subject under identical conditions. For each individual of 10 healthy intact subjects, 10 sessions were recorded in the morning (AM) and afternoon (PM) on five days, where each session involved 12 repetitions of seven hand grasps. Furthermore, each subject was required to sit in front of a table with the forearm leaning on it to grasp an object for approximately four seconds in one repetition and rest for approximately four seconds afterward. The detailed illustration of these hand grasps with different objects can be found in Figure 26. Muscle activity was measured with 14 Delsys Trigno Wireless electrodes at a sampling frequency of 2 kHz, attached to the upper half of the forearm in two rows with equal space.

This section outlines two key findings from the summary of the research work on NinaPro DB6, which is presented in Table 4. One is that NinaPro DB6 has been underexplored in the literature due to the difficulty of improving its recognition accuracy, as evidenced by the observation of generally poor recognition accuracy. As evaluated by (Chang, Phinyomark and Scheme 2020), the signal quality of this data set was acceptable. Despite the fact that some signals suffer from low Signal-to-Noise Ratio (SNR) values and incorrect labelling, we can argue that poor accuracy was due to highly confused hand grasps along with large and various variabilities. In addition to the variability between steady and transient states and the temporal variability, at least two more variabilities must be taken into account in NinaPro DB6 namely the variability between data acquisitions and the variability between objects for each hand grasp. Note that electrodes were not required to be attached at the exact same position on each data acquisition, and two objects were used in turn when each individual performed the same hand grasp. Most studies reported intra-session accuracy in NinaPro DB6 to validate their proposed algorithms. Our argument is that the primary focus when working with the NinaPro DB6 dataset should be on enhancing inter-session recognition accuracy rather than intra-session accuracy.

Table 4: Summary of the research work on NinaPro DB6 in the literature.

Paper	Model	Window+Step	Features	Inter-subject	Accuracy (%)	
					Intra-session	Inter-session within-day
Palermo et al.	RF	200ms+10ms	WL	-	52.40 ^a	25.40 ^b
Wei et al.	MV-CNN ¹	200ms+10ms	Raw sEMG ¹	-	62.90 ^c	-
Cene et al.	R-RELM ²	200ms+10ms	MAV+WL	-	69.83 ^a	41.75 ^b
	GAN			-	66.10 ^c	-
Hu et al.	GengNet	200ms+N.A.	sEMG images ³	-	56.40 ^c	-
	DuNet			-	56.80 ^c	-
	HuNet			-	58.00 ^c	-
Zanghieri et al.	TCN	150ms+15ms	Raw sEMG	-	-	49.60 ^d
Marano et al.	RF	200ms+10ms	ave(mDWT, MAV, VAR) ⁴	44.00 ^e	-	-
				75.40 ^f	75.80 ^g	-
Bao et al.	CNN	150ms+25ms	FFT ⁵	-	-	33.33 / 73.33 ^h

¹ Even though the raw sEMG was claimed to be the input of Multi-View Convolutional Neural Network (MV-CNN), each feature set would be computed as a view of the sEMG. This means that the input of MV-CNN is, in fact, the extracted features of the raw sEMG. Note that the sampling frequency has been down-sampled to 100 Hz in this work.

² R-RELM: Reliable-Regularized Extreme Learning Machines

³ The sEMG images were computed using three traditional feature sets.

⁴ The input of the model was the average of the marginal Discrete Wavelet Transform (mDWT), the Mean Absolute Value (MAV), and the Variance (VAR). Note that only the first eight electrodes were used in this work.

⁵ The Fast Fourier Transform (FFT) was applied to compute the spectrum of sEMG.

^{a, b} All odd trials collected on the morning of each day were used to train the model for each individual, which was tested on all even trials of the same morning data^a and the same afternoon data^b.

^c All odd trials collected in the morning and afternoon of five days were used to train the model for each individual, and the accuracy reported was calculated intrasession (and intrasubject) for the remaining trials.

^d The incremental training protocol was used and this result was not purely cross-day since session five, as part of the training data, was recorded on the morning of the 3rd day, while session six, as part of the testing data, was recorded on the afternoon of the same day.

^e This is the inter-subject baseline accuracy achieved with 'Source only' method, i.e., no data from the testing subject was used.

^f When allowing the use of eight repetitions of the target subject to train the model, the best recognition accuracy was reported as 75.4%.

^g Here, k -fold cross-validation was used, where k is the number of repetitions from other sessions of the same subject.

^h The CNN was trained using the data for day 1 and tested on the data for day 3 – 5 for each individual. Higher recognition accuracy was reported under the rejection scheme.

More importantly, Table 4 shows that the previous SoA inter-session accuracy was achieved as 49.6% using a proposed Temporal Convolutional Network (TCN) with multi-session training (Zanghieri et al. 2020a). Recently, Bao et al. proposed a new confidence estimate used for a CNN to improve the long-term robustness of recognition and reported an SoA inter-session between-day accuracy of 73.33% under the rejection scheme.

5.3 Multidimensional uncertainty-aware models

The proposed multidimensional uncertainty-aware models are shown in Figure 14. We first proposed simple but efficient 2D CNN and 3D CNN for sEMG-based hand grasp recognition as baseline models in this project. Based on these model structures, 2D ECNN and 3D ECNN were constructed to improve the long-term robustness of sEMG-based HGR. In this work, two dimensions of uncertainty were mainly considered: vacuity and dissonance. Recall that the uncertainty of vacuity (or known as belief vacuity) is due to insufficient or unreliable information received from sources, while the uncertainty of dissonance (or known as belief dissonance) reflects the situation where a model holds simultaneous contradicting beliefs about a given prediction, which is usually caused by valid but conflicting evidence derived from the model output (Jøsang, Cho and Chen 2018).

5.3.1 Baseline model

Taking into account the trade-off between model efficiency and training computation cost, we first proposed a baseline model that was based on a previously proposed enhanced ConvNet (Côté-Allard et al. 2019) by adding a third convolutional layer and two more fully connected layers for high-level reasoning. We refer to this baseline model simply as 2D CNN. To further optimise the efficiency

of the model, a second baseline model was designed by integrating 3D convolutional layers and presenting 14 channels of the sEMG as a matrix format of 2×7 to learn characteristics taking into account both spatial and temporal information. We refer to the second baseline model simply as 3D CNN. The intuition behind this design is to represent input sEMG signals by considering the physical information of electrode placement. It is important to note that the electrodes are arranged in two rows, and using a 3D CNN structure can help approximate the channel relationship as the kernel moves through the data. If a more precise representation of the channel relationship is required, a Graph Convolutional Network (GCN) could be a potential approach, which will be discussed in the future work, seen in subsection 7.2.2, and the detailed information of electrode placement for NinaPro DB6 is presented in Figure 23. Note that the classifiers in this task are required to output predictive probabilities of 8 classes (including the rest posture). The size of each frame of the raw sEMG signals for 2D CNN and 3D CNN is 14×400 and $400 \times 2 \times 7$, respectively.

5.3.2 Uncertainty-aware models

We constructed uncertainty-aware models by integrating two baseline models with EDL (Sensoy, Kaplan and Kandemir 2018). For the sake of consistency, they will be referred to as 2D ECNN and 3D ECNN, respectively. Initially, EDL was proposed based on the framework of Subjective Logic (SL) to help explicitly train a model that can make a prediction along with the quantified uncertainty of it, i.e., vacuity in the context of SL, indicating whether there is sufficient evidence to support model predictions (Sensoy, Kaplan and Kandemir 2018). Due to the SL framework, ECNN is capable of inferring dissonance, which is another type of explainable uncertainty derived from conflicting evidence (Jøsang, Cho and Chen 2018). They can be regarded as evidential uncertainty, and more details can be found in section 3.7.

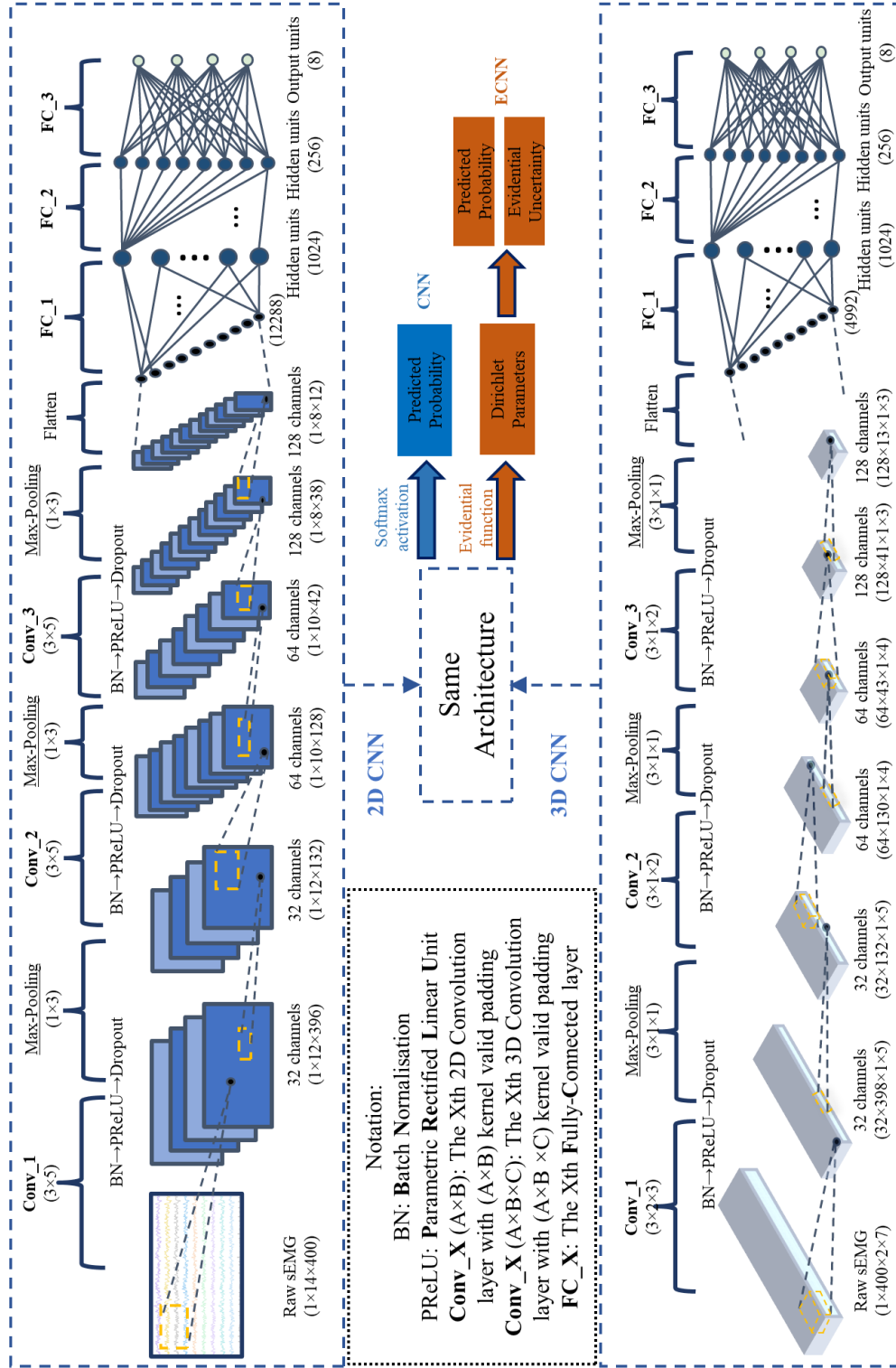


Figure 14: The model structure of Evidential Convolutional Networks (ECNN) and its conventional version (i.e., CNN).

5.4 Experiments

The versions of PyTorch and Python used in this study are 1.10.2 and 3.9.7, respectively. The experimental sequences were constructed by data loading, data segmentation, hyperparameter determination, model training, and model testing. Due to the stochastic nature of deep learning algorithms, all models were trained 10 times with the determined hyperparameters. The source code for this study is available on GitHub (<https://github.com/YuzhouLin/3dECNN-NinaProDB6>).

5.4.1 Data Preprocessing

To satisfy the constraints of real-time control of prosthetic limbs, a sliding window of 250 ms ($< 300ms$ (Hudgins, Parker and Scott 1993)) was selected with an increment of 25ms. The overlap was set to as high as 90% to increase the decision density and reduce the prediction delay. For fair consideration, the rest movement was treated identically to that of the other hand grasps. Specifically, each trial of ‘rest’ was taken after the first 0.5s to avoid incorrect labelling and only about 1/7 duration of its previous hand grasp to guarantee a balance in data for all classes. Note that the signals in NinaPro DB6 were pre-processed data that had been filtered by a bandpass filter (20 – 450Hz) and a notch filter (50 Hz). Therefore, no additional signal pre-processing was required, and each segmented sEMG would be taken as a model input directly. Furthermore, the sEMG signals used in training and testing included both steady and transient states.

Since the main objective of this study was to investigate the long-term performance of sEMG-based hand grasp recognition, the data collected on day 1 and day 2 for each individual were used as training and validation sets, while the rest were used as test sets. To capture more generalised features, the last two cycles of each acquisition were used as validation sets. Specifically, the 11th and 12th cycles recorded at AM and PM of the first two days were used as validation sets. The

ratio of training data to validation data was 5 : 1. For consistency, the 2nd and 9th subjects were not be studied in this project, and the reasons are listed below.

- Only two labels of sEMG signals were found instead of 8 on the morning of day 2 for the subject 2. This may be caused by label noise or a mistake in data acquisition.
- Only 13 valid channels were found instead of 14 on the morning of day 1 for the subject 9. Specifically, you can only observe 0 on channel 7 (that is, the 8th channel).

5.4.2 Model training

To reduce the computational burden, we used Tree-structured Parzen Estimator (TPE) (Bergstra et al. 2011; Bergstra, Yamins and Cox 2013), which is one of the SoA Hyperparameter Optimisation (HPO) algorithms, to search for optimised hyperparameters. Taking advantage of sequential model-based global optimisation algorithms (Hutter, Hoos and Leyton-Brown 2011; Bergstra et al. 2011), the available values of each hyperparameter can be determined based on the previous search results because the hyperparameters have been organised into a tree-like space. The hyperparameters for each model were determined using Optuna (Akiba et al. 2019), which is a powerful HPO framework. Each HPO search stopped after running five study trials, where each trial refers to each evaluation of an objective function to minimise validation loss. Additionally, early stop was used to avoid overfitting in each training. When no improvement was found in the validation set after waiting for 10 epochs or when the training epoch reached 200, the model stopped training. Furthermore, we use Adaptive Moment Estimation (ADAM) (Kingma and Ba 2015) as an optimiser for training and the ReduceL-RonPlateau schedule with patience for five epochs and the decay factor of 0.8 to decrease the learning rate during training.

Table 5: The Determined Optimal Hyperparameters of 2D CNN and 2D ECNNs on *NinaPro DB6*

Models	HPs	Subjects									
		1	3	4	5	6	7	8	10		
CNN	batch size	32	128	32	128	64	128	32	64		
	learning rate	0.0001692	0.000292	0.000731	0.000729	0.000201	0.001428	0.000286	0.000253		
ECNN-A	batch size	256	32	256	128	128	256	1024	64		
	learning rate	0.000257	0.000225	0.001278	0.000069	0.001503	0.000137	0.000624	0.000233		
ECNN-B	evidence func	<i>SoftPlus</i>	<i>SoftPlus</i>	<i>SoftPlus</i>	<i>SoftPlus</i>	<i>ReLU</i>	<i>SoftPlus</i>	<i>ReLU</i>	<i>SoftPlus</i>	<i>ReLU</i>	<i>SoftPlus</i>
	batch size	64	64	32	32	128	1024	32	256		
ECNN-C	learning rate	0.000149	0.000896	0.000016	0.000073	0.000128	0.001145	0.000106	0.000306		
	evidence func	<i>Exp</i>	<i>Exp</i>	<i>ReLU</i>	<i>SoftPlus</i>	<i>ReLU</i>	<i>SoftPlus</i>	<i>SoftPlus</i>	<i>ReLU</i>	<i>ReLU</i>	<i>ReLU</i>
ECNN-C	annealing step	60	35	60	15	10	60	50	45		
	batch size	128	512	512	1024	128	64	256	512		
ECNN-C	learning rate	0.000014	0.000125	0.001115	0.000341	0.000391	0.000817	0.001339	0.000388		
	evidence func	<i>ReLU</i>	<i>Exp</i>	<i>Exp</i>	<i>ReLU</i>	<i>SoftPlus</i>	<i>ReLU</i>	<i>Exp</i>	<i>ReLU</i>	<i>ReLU</i>	<i>ReLU</i>
ECNN-C	lambda	0.014773	0.018905	0.022696	0.022697	0.046129	0.010497	0.015631	0.056445		

Table 6: The Determined Optimal Hyperparameters of 3D CNN and 3D ECNNs on *NinaPro DB6*

Models	HPs	Subjects									
		1	3	4	5	6	7	8	10		
CNN	batch size	64	256	64	64	256	128	256	32	1024	32
	learning rate	0.000534	0.000807	0.000160	0.000633	0.000515	0.000306	0.000628	0.000327		
ECNN-A	batch size	64	32	1024	32	1024	32	32	1024	32	1024
	learning rate	0.001054	0.000119	0.000187	0.000035	0.000648	0.000089	0.000046	0.000435		
	evidence func	<i>ReLU</i>	<i>SoftPlus</i>	<i>ReLU</i>	<i>ReLU</i>	<i>ReLU</i>	<i>ReLU</i>	<i>ReLU</i>	<i>ReLU</i>	<i>ReLU</i>	<i>SoftPlus</i>
ECNN-B	batch size	128	1024	32	32	256	1024	1024	128		
	learning rate	0.000781	0.003465	0.000543	0.000271	0.000202	0.000387	0.000387	0.001061		
	evidence func	<i>Exp</i>	<i>SoftPlus</i>	<i>SoftPlus</i>	<i>SoftPlus</i>	<i>ReLU</i>	<i>ReLU</i>	<i>ReLU</i>	<i>ReLU</i>	<i>ReLU</i>	<i>SoftPlus</i>
	annealing step	35	40	50	25	50	15	55	60		
ECNN-C	batch size	512	256	1024	32	512	1024	64	512		
	learning rate	0.001765	0.000852	0.000556	0.000030	0.000138	0.000624	0.000393	0.001981		
	evidence func	<i>ReLU</i>	<i>Exp</i>	<i>ReLU</i>	<i>SoftPlus</i>	<i>SoftPlus</i>	<i>ReLU</i>	<i>SoftPlus</i>	<i>SoftPlus</i>	<i>ReLU</i>	<i>SoftPlus</i>
	lambda	0.060979	0.152230	0.011926	0.039430	0.074864	0.030441	0.011291	0.092692		

Eventually, the results of the search for the hyperparameters determined of all the models used for each individual can be seen in Table 5 and Table 6.

5.4.3 Rejection-capable performance evaluation

Assume there is an uncertainty threshold δ , a prediction can be considered uncertain when its quantified uncertainty is greater than or equal to δ and certain otherwise. By determining δ , a model is capable of making only confident predictions by rejecting uncertain ones. When $\delta = 1$, it simply refers to standard recognition since no rejections will be made. Here, the rejection threshold of a model for each individual in each run was determined by validating its performance in misclassification detection in the validation set. In more detail, a rejection threshold was selected when it achieved the best F_β -Score,

$$F_\beta\text{-Score} = (1 + \beta^2) \frac{\text{Precision} \cdot \text{Recall}}{(\beta^2 \cdot \text{Precision}) + \text{Recall}}, \quad (24)$$

where β was chosen as 2 in this study, which means that recall was considered 2 times as important as precision. As shown in Figure 15, the purpose of this setting was to maximise True Acceptance Rate (TAR) as much as possible in testing data sets with a determined rejection threshold. When the number of True Positive (TP) and True Negative (TN) remains the same, a higher recall can only be achieved with fewer False Negatives (FNs), thus indirectly producing a higher TAR. Furthermore, to avoid setting extremely low thresholds so that no active predictions would be made in the testing phase, the determined threshold had to be checked first on the validation sets. If the number of active predictions was less than the 10% of the total number of validation samples in each acquisition, the threshold was selected again on the value which has the second best F_β -Score and et cetera.

In all experiments, the results would be reported as an average of 10 runs,

		Predicted Condition		
		Positive (Reject)	Negative (Accept)	
Actual Condition	Positive (Wrong)	True Positive (TP)	False Negative (FN)	Recall $\frac{TP}{TP+FN} = \frac{\text{[TP]}}{\text{[TP+FN]}}$
	Negative (Correct)	False Positive (FP)	True Negative (TN)	
		Precision $\frac{TP}{TP+FP} = \frac{\text{[TP]}}{\text{[TP+FP]}}$	True Acceptance Rate (TAR) $\frac{TN}{TN+FN} = \frac{\text{[TN]}}{\text{[TN+FN]}}$	Accuracy $\frac{FP+TN}{TP+FN+FP+TN}$

Figure 15: The confusion matrix of misclassification detection. TAR can be considered the accuracy of rejection-capable sEMG-based hand grasp recognition and Precision is actually the True Rejection Rate (TRR) defined in this chapter.

and the Wilcoxon signed rank test was used to compare the performance of CNN (baseline model) and ECNN variants. There was a significant difference in the results between the two models when the p-value < 0.05, that is, the null hypothesis, which assumed that two related paired samples come from the same distribution, was rejected. The calculations of all the evaluation metrics used in this section can be found in Figure 15.

5.4.4 Recognition Performance - No rejections

To investigate the long-term robustness of sEMG-based hand grasp recognition with the proposed uncertainty-aware models, we first present their recognition

Table 7: Inter-Session Cross-Day Standard Recognition Accuracy of ConvNets with Comparisons

Models	Day3	Day4	Day5	Ave.
2D CNN	52.5±11.8	54.7±11.2	48.1±13.5	51.8±12.5
2D ECNN-A	53.1±11.4*	55.2±10.5*	49.0±12.6	52.4±11.8
2D ECNN-B	50.5±10.8	53.6±08.9	47.7±11.6	50.6±10.8
2D ECNN-C	52.0±11.9	55.2±10.1*	47.7±13.0	51.6±12.1
3D CNN	52.0±13.5	52.2±11.6	48.3±11.7	50.8±12.3
3D ECNN-A	52.4±12.8*	53.1±10.1*	48.8±11.4*	51.4±11.6
3D ECNN-B	47.4±10.2	49.2±08.5	44.3±09.7	47.0±09.5
3D ECNN-C	51.0±13.2	51.8±10.7	47.4±11.7	50.1±11.9

¹ All results were reported as the mean and standard deviation of classification accuracy for 10 runs over 8 subjects.

² The Wilcoxon signed rank test is applied to compare CNN with the ECNN variants, and the null hypothesis (H0) assumes that the median of the differences between two distributions is zero, which will be rejected when the p-value < 0.05. For conciseness, the asterisk (*) is used when H0 is not rejected here.

accuracy when there was no rejection option. Recall that the proposed CNN was considered a baseline model so that its performance could be compared with the ECNN variants. As shown in Table 7, the recognition accuracy of all 2D models increases slightly (about 5.1%) on day 4 and decreases greatly (about 12%) on day 5. Similarly, 3D models present the same performance, but with fewer changes, i.e., their average recognition accuracy increases slightly (only about 1.7%) on day 4 and decreases greatly (only about 6.9%) on day 5. In terms of 2D models, the best average accuracy was observed in the 2D ECNN-A, which outperformed 2D CNN with a small but statistically significant difference of 0.67%. Again, 3D ECNN-A also achieved the best accuracy among all 3D models, outperforming 3D CNN with a small but statistically significant difference of 0.6%.

5.4.5 Recognition Performance - With Rejections

To better investigate the rejection-capable performance of sEMG-based hand grasp recognition with respect to all proposed models, we employed three evaluation metrics, including the TAR, the TRR, and the Rejection Rate (RR). The rejection-capable performances of models with different uncertainty scores are summarised in Table 8 and Table 9. Recall that TAR can be considered the recognition accuracy under the rejection scheme.

First, the same commonly seen uncertainty estimate was used for a fair comparison between the 2D/3D CNN and 2D/3D ECNN variants. It can be seen that both 2D and 3D ECNN variants significantly outperformed 2D and 3D CNN with $u_{nEntropy}$ or u_{nnmp} by presenting higher TARs. In terms of 2D models, 2D ECNN-B achieved the highest TAR of 77.77%, which was 3.57% higher than 2D CNN when using u_{nnmp} as an uncertainty estimate. Correspondingly, 3D ECNN-B also achieved the highest TAR of 76.77%, which was 7.07% higher than 3D CNN when using u_{nnmp} as an uncertainty estimate. Taking into account the use of evidential uncertainty, both 2D and 3D ECNN-B achieved the best performance with u_{vac} compared to other 2D and 3D variants of ECNN.

Furthermore, it was found that the difference between 2D/3D ECNN-B and 2D/3D ECNN-A in TAR was not statistically significant but either 2D or 3D ECNN-B made fewer incorrect rejections, as evidenced by the observation of a 5.74%/8.56% higher TRR and a 4.81%/4.35% lower RR. When using u_{diss} as the uncertainty score, it is important to note that both 2D and 3D ECNN-B achieved the lowest RR, which was only about 50%. With more predictions accepted, they produced the lowest but acceptable TAR of 68.31% and 65.83%.

Table 8: Overall Rejection-capable Performance Comparison of 2D ConvNets

Scores	Models	True Acceptance Rate \uparrow		True Rejection Rate \uparrow		Rejection Rate \downarrow	
		M(%) \pm SD(%)	H0 (p)	M(%) \pm SD(%)	H0 (p)	M(%) \pm SD(%)	H0 (p)
$u_nEntropy$	CNN	74.82 \pm 14.80	-	61.69 \pm 9.29	-	62.74 \pm 10.89	-
	ECNN-A	76.23 \pm 13.03	0 (1.1e-06)	60.61 \pm 9.74	0 (5.3e-14)	64.27 \pm 9.74	0 (4.4e-13)
	ECNN-B	77.76 \pm 12.81	0 (1.7e-18)	63.70 \pm 7.78	0 (3.1e-31)	65.62 \pm 12.49	0 (4.0e-19)
	ECNN-C	76.31 \pm 14.07	0 (1.3e-11)	61.55 \pm 9.97	1 (4.3e-01)	64.84 \pm 10.18	0 (2.3e-19)
u_{nmmmp}	CNN	74.20 \pm 14.82	-	61.85 \pm 9.08	-	61.89 \pm 10.94	-
	ECNN-A	75.72 \pm 13.48	0 (2.7e-09)	60.60 \pm 9.58	0 (1.9e-21)	63.72 \pm 9.62	0 (2.6e-18)
	ECNN-B	77.77 \pm 12.82	0 (4.2e-28)	63.71 \pm 7.74	0 (6.3e-29)	65.62 \pm 12.54	0 (1.0e-34)
	ECNN-C	76.26 \pm 14.15	0 (5.3e-22)	61.46 \pm 9.99	0 (3.7e-03)	64.93 \pm 10.04	0 (3.4e-34)
u_{vac}	ECNN-A	77.92 \pm 12.31	-	57.86 \pm 11.04	-	70.55 \pm 11.18	-
	ECNN-B	77.86 \pm 12.88	1 (5.7e-01)	63.60 \pm 7.85	0 (7.7e-83)	65.74 \pm 12.63	0 (5.6e-50)
	ECNN-C	70.00 \pm 19.39	0 (10.0e-49)	60.84 \pm 10.20	0 (1.1e-53)	62.06 \pm 16.96	0 (1.1e-56)
u_{diss}	ECNN-A	74.14 \pm 14.43	-	59.25 \pm 9.39	-	64.65 \pm 8.48	-
	ECNN-B	68.31 \pm 20.02	0 (7.8e-13)	58.12 \pm 10.91	0 (1.0e-04)	51.75 \pm 33.64	0 (5.7e-11)
	ECNN-C	71.61 \pm 14.16	0 (7.5e-20)	61.48 \pm 9.65	0 (4.3e-61)	59.76 \pm 12.03	0 (2.8e-29)
u_{vac} & u_{diss}	ECNN-A	83.00 \pm 12.11	-	56.29 \pm 10.63	-	77.46 \pm 8.79	-
	ECNN-B	81.04 \pm 10.54	0 (2.0e-05)	59.94 \pm 10.91	0 (1.5e-51)	74.32 \pm 12.16	0 (1.1e-12)
	ECNN-C	70.67 \pm 29.22	0 (2.4e-82)	58.51 \pm 10.21	0 (6.8e-62)	73.42 \pm 12.12	0 (5.8e-55)

¹ An upward arrow indicates that a higher value is desirable and vice versa.

² The M and SD refer to the mean and standard deviation of each evaluation metric for 10 runs over 8 subjects.

³ The Wilcoxon signed rank test is applied to compare CNN with ECNN variants with commonly used uncertainty measures and to compare ECNN-A with other ECNN variants with evidential uncertainty measures. The null hypothesis (H0) assumes that the median of differences between two distributions is zero, which will be rejected when the p-value < 0.05 .

* This multidimensional uncertainties-based rejection performance is achieved by rejecting predictions with high vacuity or dissonance. Specifically, a prediction will be accepted only when its evidential uncertainties are less than predetermined thresholds, i.e. $u_{vac} < \delta_{vac}$ and $u_{diss} < \delta_{diss}$.

Table 9: Overall Rejection-Capable Performance Comparison of 3D ConvNets

Scores	Models	True Acceptance Rate \uparrow		True Rejection Rate \uparrow		Rejection Rate \downarrow	
		M(%) \pm SD(%)	H0 (p)	M(%) \pm SD(%)	H0 (p)	M(%) \pm SD(%)	H0 (p)
$u_nEntropy$	CNN	72.28 \pm 15.62	-	62.81 \pm 9.28	-	60.81 \pm 9.57	-
	ECNN-A	75.50 \pm 14.75	0 (3.7e-26)	61.64 \pm 9.19	0 (5.7e-09)	64.62 \pm 8.55	0 (2.6e-30)
	ECNN-B	76.71 \pm 15.20	0 (1.1e-13)	65.83 \pm 7.48	0 (5.1e-19)	69.23 \pm 10.80	0 (7.6e-53)
	ECNN-C	73.84 \pm 16.44	0 (1.5e-07)	63.84 \pm 9.97	0 (2.6e-07)	62.55 \pm 9.50	0 (1.0e-08)
u_{nmmmp}	CNN	71.70 \pm 15.60	-	62.95 \pm 9.09	-	60.00 \pm 9.68	-
	ECNN-A	75.49 \pm 14.63	0 (7.3e-31)	61.47 \pm 9.12	0 (5.6e-14)	64.90 \pm 8.80	0 (2.6e-39)
	ECNN-B	76.77 \pm 15.17	0 (1.4e-18)	65.84 \pm 7.46	0 (2.7e-18)	69.28 \pm 10.75	0 (1.2e-58)
	ECNN-C	73.82 \pm 16.38	0 (1.3e-10)	63.76 \pm 9.85	0 (4.5e-05)	62.77 \pm 9.35	0 (6.9e-14)
u_{vac}	ECNN-A	76.11 \pm 15.13	-	57.23 \pm 11.66	-	73.63 \pm 10.26	-
	ECNN-B	76.63 \pm 15.13	1 (8.7e-01)	65.79 \pm 7.46	0 (3.1e-50)	69.28 \pm 10.86	0 (1.3e-15)
	ECNN-C	73.77 \pm 17.18	0 (1.8e-09)	63.26 \pm 11.31	0 (1.4e-68)	63.03 \pm 11.42	0 (3.9e-59)
u_{diss}	ECNN-A	75.06 \pm 14.71	-	59.83 \pm 9.59	-	67.70 \pm 9.07	-
	ECNN-B	65.83 \pm 20.22	0 (6.7e-26)	60.98 \pm 9.93	0 (9.4e-05)	52.56 \pm 32.27	0 (5.3e-10)
	ECNN-C	70.96 \pm 15.52	0 (1.4e-24)	63.01 \pm 9.66	0 (4.8e-45)	61.22 \pm 12.02	0 (1.8e-31)
u_{vac} & u_{diss}	ECNN-A	83.51 \pm 15.43	-	55.88 \pm 11.35	-	81.43 \pm 7.65	-
	ECNN-B	80.32 \pm 16.00	0 (1.8e-13)	62.82 \pm 9.79	0 (7.3e-44)	76.64 \pm 10.20	0 (1.7e-17)
	ECNN-C	77.35 \pm 16.53	0 (9.8e-57)	61.30 \pm 10.61	0 (1.1e-69)	70.23 \pm 9.08	0 (2.9e-78)

¹ An upward arrow indicates that a higher value is desirable and vice versa.

² The M and SD refer to the mean and standard deviation of each evaluation metric for 10 runs over 8 subjects.

³ The Wilcoxon signed rank test is applied to compare CNN with ECNN variants with commonly used uncertainty measures and to compare ECNN-A with other ECNN variants with evidential uncertainty measures. The null hypothesis (H0) assumes that the median of differences between two distributions is zero, which will be rejected when the p-value < 0.05 .

* This multidimensional uncertainties-based rejection performance is achieved by rejecting predictions with high vacuity or dissonance. Specifically, a prediction will be accepted only when its evidential uncertainties are less than predetermined thresholds, i.e., $u_{vac} < \delta_{vac}$ and $u_{diss} < \delta_{diss}$.

Special attention was paid to the use of multidimensional uncertainties (that is, predictions were accepted when their u_{vac} and u_{diss} were found to be smaller than their respective predetermined thresholds simultaneously). Under this condition, 2D ECNN-A achieved the highest TAR (83%) while 2D ECNN-B obtained the best comprehensive performance by presenting a lower TAR (81.04%) but with 3.14% more active predictions. Correspondingly, 3D ECNN-A also achieved the highest TAR (83.51%) while the best comprehensive performance was achieved by 3D ECNN-B since it presented a lower TAR (80.32%) but with 4.79% more active predictions.

To investigate the repeatability of sEMG-based hand grasp recognition with the proposed ECNN models with respect to multidimensional uncertainty, the grouped violin plot was drawn to visualise their differences in TAR in several data acquisitions over days. A violin plot (Hintze and Nelson 1998) is a hybrid of a box plot and a kernel density plot, which can be used to compare the distributions of several groups. As shown in Figure 16, the median difference in TAR between acquisitions for each 2D/3D ECNN variant was not pronounced until the afternoon of day 5. Furthermore, in terms of 2D models, both ECNN-A and ECNN-B obtained robust long-term sEMG-based hand grasp recognition performance, and ECNN-A was slightly more robust than ECNN-B from observation of its fewer position shifts of wider sections as time passes. This finding could also be observed in 3D ECNN models. Note that a wider section of the violin plot represents a higher probability that a model will appear on the given TAR. To further investigate the rejection-capable sEMG-based hand grasp recognition performance of ECNN-A and ECNN-B on each class, the confusion matrix results of ECNN-A and ECNN-B were plotted, where the 2D versions can be seen in Figure 17 and 3D versions in Figure 18. Both of them will be discussed later in section 5.5.

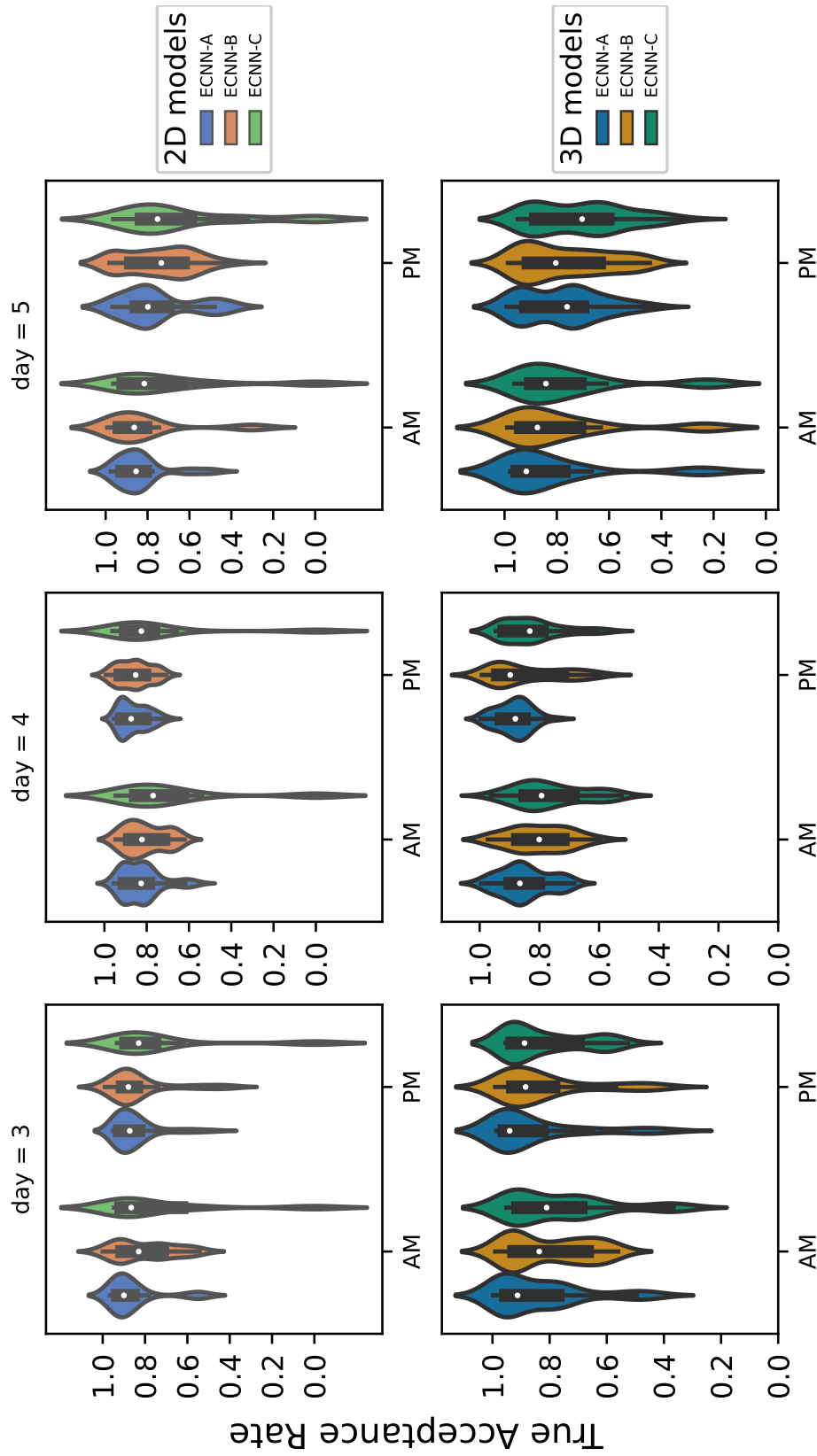


Figure 16: Multidimensional uncertainty-based rejection-capable accuracy in several data acquisitions across days

Table 10: Comparison of the proposed models with the SoA results on the inter-session cross-day accuracy

Model	Window + Step	Features	Inter-session cross-day accuracy (%)	
			Without rejection	With Rejection
TCN (Zanghieri et al. 2020a)	150ms + 15ms	raw sEMG	49.6 ¹	-
CNN (Bao et al. 2022)	150ms + 25ms	the spectrum of sEMG	33.33	73.33 ²
2D ECNN-A	200ms + 20ms	raw sEMG	52.44	83.00 *
2D ECNN-B	200ms + 20ms	raw sEMG	50.61	81.04 *
2D ECNN-C	200ms + 20ms	raw sEMG	51.63	76.31 **
3D ECNN-A	200ms + 20ms	raw sEMG	51.44	83.51 *
3D ECNN-B	200ms + 20ms	raw sEMG	46.97	80.32 *
3D ECNN-C	200ms + 20ms	raw sEMG	50.07	77.35 *

¹ The incremental training protocol was used and this result was not purely cross-day, since the session 5, as part of the training data, was recorded on the morning of day 3, while the session 6, as part of the testing data, was recorded on the afternoon of the same day. Hence, assuming the strict inter-session cross-day training protocol was used, the actual accuracy would be lower than the reported one.

² This accuracy was calculated by 1 – its proposed error rate, which only considers the number of false acceptance samples. Hence, the actual accuracy would be lower than their reported one.

* This best rejection-capable accuracy was reported based on multidimensional uncertainties, i.e., u_{vac} and u_{diss} .

** This best rejection-capable accuracy was reported based on $u_{nEntropy}$.

5.4.6 Comparison with SoA

The comparison of the inter-session cross-day recognition accuracy in terms of non-rejection and rejection with the previous study was presented in Table 10. It is observed that our proposed uncertainty-aware models outperformed SoA models under all conditions. Under the non-rejection scheme, 2D ECNN-A achieved the best performance, which significantly improves the previous SoA recognition accuracy by 5.72%. When allowing rejecting uncertain predictions, the best performance was achieved by 3D ECNN-A, which significantly improves the previous SoA recognition accuracy by 13.88%.

5.5 Discussion

NinaPro DB6 as a valuable benchmark dataset deserves more research because it was built to be challenging by including variability from data acquisitions over days and aimed at improving the long-term robustness of hand prostheses control systems. It is worth noting that electrodes were not required to be placed in expected locations exactly during each data acquisition, which approaches real-life scenarios and brings the challenge of remaining the long-term robustness of sEMG-based hand grasp recognition to the table. In this study, we proposed uncertainty-aware models (2D and 3D ECNNs) to address this challenge by rejecting doubtful predictions with multidimensional uncertainties. The potential of ECNN has been explored with three training strategies implemented by slightly different loss functions. The main implication of our results is that a CNN can be easily modified as an uncertainty-aware model by integrating with EDL. Leaving aside the training strategies used by ECNN, all ECNN variants statistically significantly outperformed CNN under the rejection scheme, even if the overall mean recognition accuracy of 2D/3D CNN was found to be 1.16%/3.80% and 0.14%/0.7% higher than those of 2D/3D ECNN-B and ECNN-C when there is no

rejection option. It should be noted that both 2D and 3D ECNN-B achieved the best TAR and TRR with a common uncertainty $u_{nEntropy}$ or u_{nnmp} , although the lowest standard accuracy were achieved by them under the non-rejection scheme. This suggests that model selection based on standard recognition accuracy is not sufficient, especially when rejection is introduced.

Table 11: Reliability Comparison of 2D and 3D ConvNets by Evaluating the Misclassification Detection regarding Uncertainty Estimates

Models	Scores			
	$u_{nEntropy}$	u_{nnmp}	u_{vac}	u_{diss}
2D CNN	54.75 ± 2.61	54.39 ± 2.80	-	-
2D ECNN-A	53.87 ± 3.09	53.40 ± 2.93	50.43 ± 4.86	45.89 ± 3.13
2D ECNN-B	57.46 ± 2.98	57.61 ± 3.04	57.24 ± 3.45	40.80 ± 7.64
2D ECNN-C	55.09 ± 3.16	54.67 ± 3.02	50.69 ± 5.40	45.89 ± 5.85
3D CNN	54.65 ± 3.44	54.38 ± 3.69	-	-
3D ECNN-A	55.29 ± 3.59	54.50 ± 3.50	50.91 ± 5.69	45.35 ± 4.46
3D ECNN-B	58.82 ± 3.06	59.88 ± 4.12	59.76 ± 4.27	40.91 ± 7.44
3D ECNN-C	56.56 ± 4.27	56.18 ± 3.95	55.61 ± 4.86	45.54 ± 5.18

Each cell refers to the reliability (R_{nAUPRC}) of a model with an uncertainty score in a format of mean and standard deviation, which were calculated for 10 runs over 8 subjects.

Following our previously proposed reliability analysis (Lin et al. 2022), the reliability of each model with an uncertainty score was obtained by using the normalised Area Under Precision-Recall Curve (nAUPRC) to indicate the performance of its misclassification detection. It can be seen in Table 11, ECNN-B achieved the best mean reliability under most conditions. In terms of 2D models, ECNN-B outperformed CNN in model reliability where the largest difference was observed as 3.22% with u_{nnmp} . Regarding the reliability comparison between it and other ECNNs with u_{vac} , 2D ECNN-B achieved the best mean reliability of 57.24%, which was 6.81% and 6.55% larger than 2D ECNN-A and ECNN-C,

respectively. In terms of 3D models, ECNN-B achieved the same performance in model reliability. These match the comparison result of rejection-capable recognition performance, indicating that ECNN-B is more reliable than others. Furthermore, it can be observed that 3D ECNN-B achieved the model reliability of 59.76% with u_{vac} , which was 2.52% higher than 2D ECNN-B. However, it did not obtain better rejection-capable recognition performance than 2D ECNN-B. This may indicate the potential of 3D ECNN-B has not been fully explored. An indirect evidence can be brought from observations of rejection-capable recognition performance with multidimensional uncertainties, presenting that 3D ECNN-B obtained the mean TAR of 80.32%, which was only 0.72% lower than 2D ECNN-B, but obtained the mean TRR of 62.82%, which was 2.88% higher than 2D ECNN-B. Under any circumstances, these findings are consistent with previous research showing that reliability analysis can be considered a useful supplementary measure for studying sEMG-based HGR.

The outstanding rejection-capable hand grasp recognition performance achieved by ECNN variants with multidimensional uncertainties confirms that evidential uncertainties are not only understandable, but also useful in practical terms. By rejecting predictions with either insufficient or conflicting evidence, ECNN variants are able to produce robust long-term hand grasp recognition. This implies that model reliability is associated with model robustness. However, it may be quite difficult to determine which model yields the best rejection-capable performance, as this is actually a trade-off problem. Ideally, if a model has a standard accuracy of 50% and knows exactly what it does not know, it will obtain 100% TAR and 100% TRR with 50% RR by an optimal rejection threshold under the rejection scheme. A comprehensive analysis is required to consider the efficiency of an uncertainty-aware model by making a comparison between the ideal and practical recognition performances. For example, 2D ECNN-B can be considered more efficient than 2D ECNN-A because it obtained an improvement of 60.16%

in recognition accuracy by making 24.92% more rejections, while 2D ECNN-A obtained an improvement of 58.40% in recognition accuracy by making 29.80% more rejections compared to the ideal situation.

One concern shown in Figure 16 was that 2D ECNN-C would achieve 0% TAR on some runs because no accepted predictions were found. This indicates that the approach used in this paper to determine the rejection threshold may sometimes yield a very strict threshold. If the results with 0% TAR and 100% RR were excluded, the mean TAR, TRR and RR obtained by 2D ECNN-C were 80.76%, 58.06% and 69.63%, respectively. This shows clearly that 2D ECNN-C can obtain better rejection-capable recognition with appropriate rejection thresholds. On the contrary, it can be said that the rejection-capable performance of a model suffers from the limitation that it is highly dependent on optimal rejection thresholds.

The confusion matrices of the predictions made by 2D ECNN-A and 2D ECNN-B under no rejection and rejection scheme can be seen in Figure 17, while the same but for 3D ECNN-A and 3D ECNN-B can be seen in Figure 18. The first finding from these figures is that it is difficult to improve the recognition accuracy of some classes, which were confused with others, by rejections. For example, one can observe that G2 and G4 were difficult to be classified correctly. When the highest number of rejections ($\geq 90\%$ RR) was made in the samples labelled G4, only a limited improvement in TAR was noticed for all models. The recognition accuracy of G4 even decreased by approximately 50% after making approximately 93% rejections in 3D ECNN-B. Associated with (i) the relatively poor rejection-capable performance achieved by the ECNN variants with u_{diss} shown in Table 8 and Table 9, (ii) poor model reliability with u_{diss} shown in Table 11, it reveals a limitation in the training of ECNN variants.

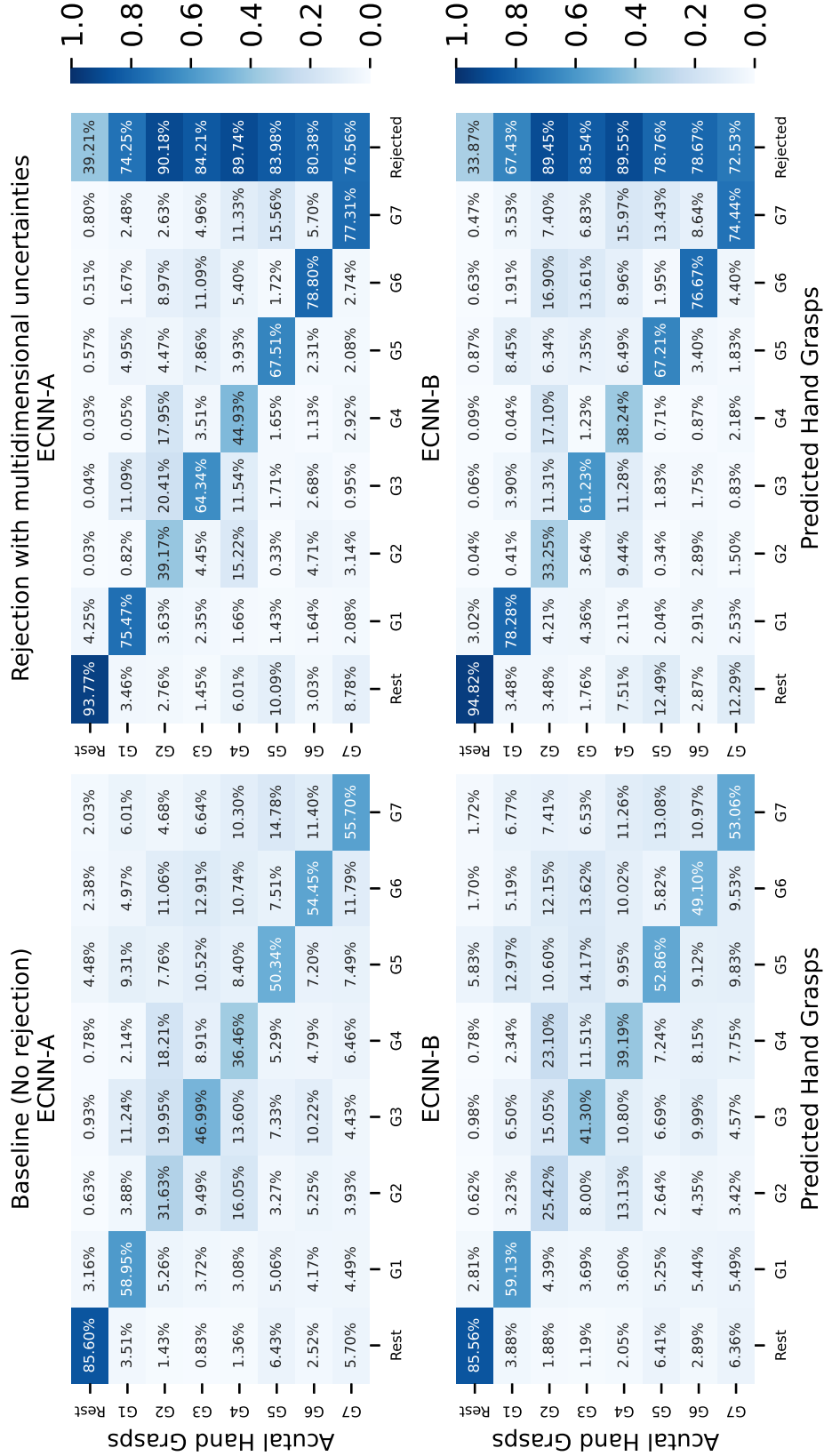


Figure 17: The normalised confusion matrices of 2D ECNN-A and 2D ECNN-B averaged for 10 runs over 8 subjects. Note that an extra x-axis was introduced when presenting confusion matrix with rejection, which should be considered a separated part from the standard one where each cell refers to the rejection rate for a class.

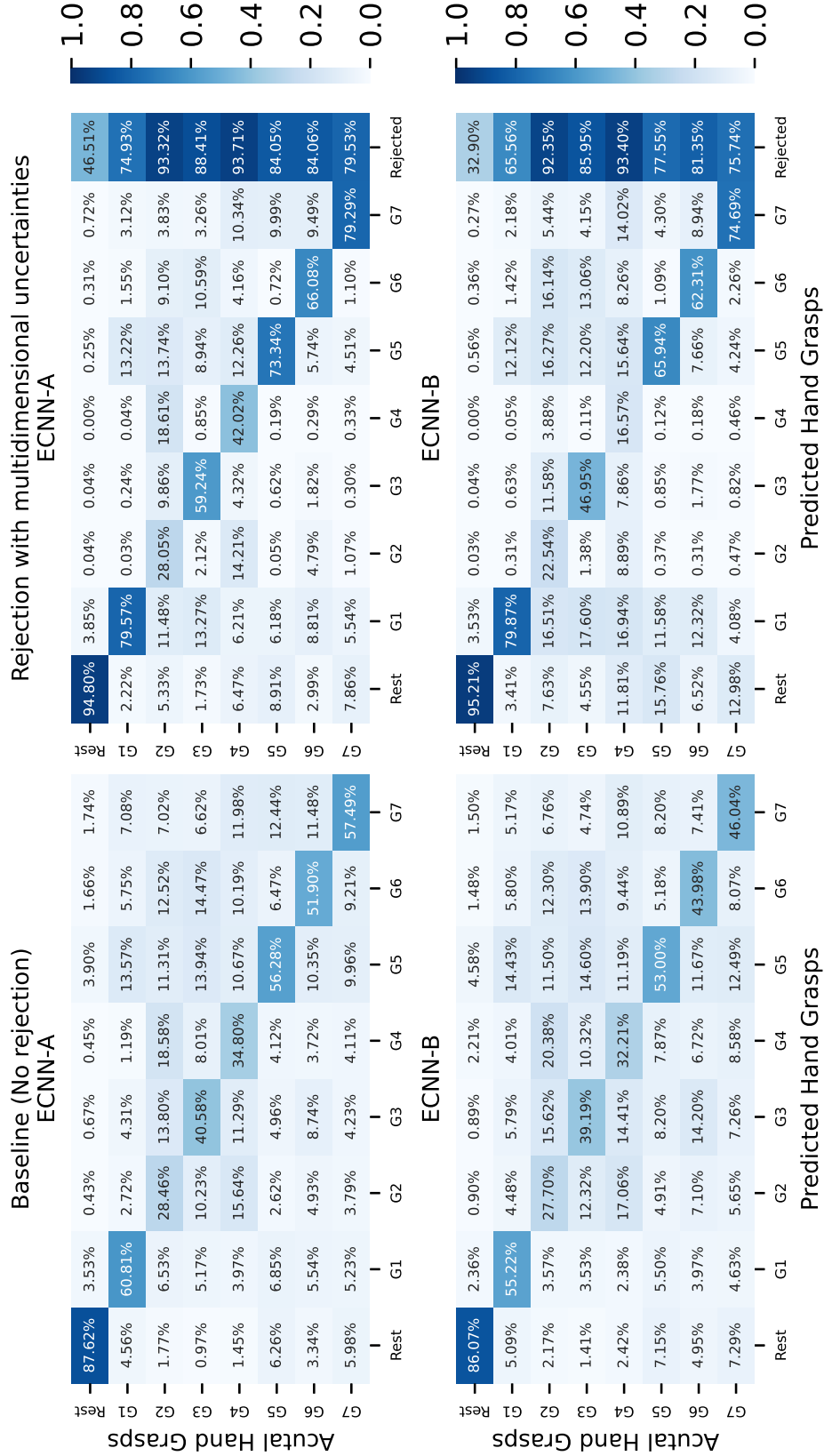


Figure 18: The normalised confusion matrices of 3D ECNN-A and 3D ECNN-B averaged for 10 runs over eight subjects. Note that an extra x-axis was introduced when presenting confusion matrix with rejection, which should be considered a separated part from the standard one where each cell refers to the rejection rate for a class.

The second finding from these confusion matrices is that there is a positive correlation between baseline performance and rejection-capable performance. For example, 3D ECNN-B obtained the highest classification accuracy on ‘rest’ as 86.07% and 95.21% with respect to the non-rejection options and with rejection. This example also shows that not surprisingly, the ‘Rest’ is considered the easiest class. Leaving the easiest and hardest classes (‘Rest’, G2 and G4) aside, it was found that about 20% improvements in hand grasps could be obtained by allowing uncertain rejections with multidimensional uncertainties regarding both 2D or 3D ECNN-A and ECNN-B. This implies that one can focus on improving the straightforward baseline performance first without suffering from the limitations of investigating its rejection-capable performance when designing an uncertainty-aware model integrated by evidential deep learning.

5.6 Conclusion

This study uses a very challenging benchmark dataset, NinaPro DB6, to demonstrate the potential of designing an uncertainty-aware model to improve the long-term robustness of hand grasp recognition with raw sEMG signals. The proposed ECNN allows us to reject predictions with high evidential uncertainty from either less supported or conflicting evidence. When there is no rejection option, both 2D and 3D ECNN outperformed the existing SoA with a significant improvement of 5.72% and 3.71% in recognition accuracy. More importantly, they achieved a high recognition accuracy of 83% and 83.51%, which were found to improve the existing SoA by 13.19% and 13.88% under the rejection scheme. Furthermore, their long-term robustness has been verified by presenting a high ($> 80\%$) rejection-capable recognition accuracy on each of 3 days with only a small degradation observed in the afternoon of day 5. This encourages us to extend the investigation to amputee subjects in improving the real-time sEMG-based control system.

Chapter 6

Improve the Inter-subject recognition with normalisation

6.1 Introduction and related work

Recently, the inter-subject Surface Electromyography (sEMG)-based Hand Gesture Recognition (HGR) has drawn more attention since the non-requirement of training data from the user is more attractive in real-life applications. However, this is a challenging research problem since sEMG signals are highly subject-specific (Matsubara, Hyon and Morimoto 2011; Halaki and Gi 2012). This means that a large variability of sEMG signals collected from different users can be observed even if they are performing the same motion in the same way with the same electrode position. In other words, the difficulty comes from the subject variability, which could also be referred to as the problem of domain shift, that is, the distribution of the target domain is different from the source domain.

To address this problem, Kerber, Puhl and Krüger used a normalisation approach (wrt the highest measured peak) to the segmented sEMG and computed

10 features from the normalised data to train a Support Vector Machine (SVM). An overall 95% accuracy was reported for classifying five hand gestures with a majority vote for an active segment. However, the user-independent classification accuracy drops to about 50% when the number of hand gestures increases to 12. Khushaba proposed a framework based on Canonical Correlation Analysis (CCA) to compute a series of style-independent features and the average classification accuracy of 82.96% with majority vote was reported to classify 12 finger movements (including resting state) by training data from normal subjects and testing with data from amputees.

Recently, many deep learning approaches, such as Convolutional Neural Network (CNN), have been proposed to improve the robustness of a classifier. Furthermore, the Transfer Learning (TL) has been widely used to solve the domain shift problem in the field of deep learning. Côté-Allard et al. proposed a ConvNet that employed raw sEMG as input and a 97.39% accuracy was obtained to classify seven hand gestures, including neutral by applying TL with four cycles of training data from the target domain where a cycle is defined as the recording of all hand gestures required to be classified. In this chapter, we propose a referencing min-max normalisation approach to re-weight the source domain sEMG data to simply reduce the domain shift. Three datasets, where differences due to many factors such as recording positions, acquisition setups, the number of hand gestures and subjects, are used to validate the inter-subject gesture recognition performance of our proposed method with Leave-One-Subject-Out Cross-Validation (LOSOCV). Finally, the result is compared to the state-of-the-art (SoA) TL approach (Côté-Allard et al. 2019).

Table 12: Description of muscle location and sEMG electrode placement in *G. Dataset*

Ch.	Muscles	Ch.	Muscle
1	Extensor digitorum(upper)	10	Extensor digitorum(lower)
2	Anconeus	3	Flexor carpi ulnaris
8	Pronator teres (upper)	4	Pronator teres (lower)
5	Flexor carpi radialis (upper)	6	Flexor carpi radialis (lower)
7	Palmaris longus	9	Extensor carpi ulnaris
11	Extensor carpi radialis brevis	12	Extensor carpi radialis longus
13	Abductor pollicis brevis	14	Abductor digiti minimi

¹ Ch. refers to sEMG channel number here.

² The description order depends on muscle location rather than channel number.

6.2 Datasets

6.2.1 *G. Dataset*

This dataset has been introduced in (Lin et al. 2018), (Li et al. 2011). Here, we refer to this dataset as *G. Dataset* since both non-invasive electrodes and amplifiers used for recording are from Guger Technologies, Graz, Austria. The data were recorded by five intact right-handed subjects with the 16-channel sEMG electrodes placed over 12 muscles located in the upper arm, forearm, and hand. A description of muscle location and sEMG electrode placement can be seen in Table 12 and more details can be found in (Lin et al. 2018). During acquisition, subjects were asked to repeat 12 hand activities including extension, flexion, pronation, supination, clench, stretch fingers, rotate hand (clockwise), rotate hand (anticlockwise), grip, thumb up, thumb and index up four times in two alternating conditions between extension (*G.0*) and relaxation (*G.1*) with the elbow resting on the arm of a chair. A subset of hand gesture illustrations can be seen in Figure 27. Here, each *cycle* includes 12 sEMG onset, which were visually determined. In the data preprocessing stage, all sEMG was first down-sampled by the scale of four, that is, 1200 Hz to reduce the processing time, but also to keep sufficient information since the usable energy of sEMG is mainly distributed from 0 to 500 Hz in terms

of frequency range (Wang, Tang and Bronlund 2013). Furthermore, a 250 ms sliding window was applied with an overlap of 225 ms ($90\% \times 250\text{ms}$) to split the sEMG onsets, allowing one to make a prediction under real-time usage constraints (≤ 300 ms (Hudgins, Parker and Scott 1993; Englehart and Hudgins 2003)).

6.2.2 *M. Dataset*

The *M. Dataset* (Côté-Allard et al. 2019) is a recent public sEMG dataset used to classify hand gestures. It contains two sub-datasets, one for pre-training and the other for evaluating. In this project, a sub-dataset called 'Test0', which is referred to as *M.0* here, is used to evaluate the performance of inter-subject sEMG-based HGR with our proposed method. The 8-channel sEMG signals were recorded with a Myo armband from 17 intact subjects making 7 hand gestures (Neutral, Hand close, Wrist extension, Ulnar deviation, Hand open, Wrist flexion, Radial deviation) for four times by standing up and having their forearm parallel to the floor. The illustration of these hand gestures can be seen in Figure 28. Here, the data were segmented by applying a sliding window of 260 ms with an overlap of 235 ms as suggested in (Côté-Allard et al. 2019).

6.3 Hypotheses

There are two hypotheses in this study:

1. The distribution of a signal from a channel
 - varies little when an individual is repeating a hand gesture many times (This may be supported by high intra-subject HGR accuracy presented in chapter 4 and chapter 5);
 - varies a lot across subjects when they are performing the same hand gesture.

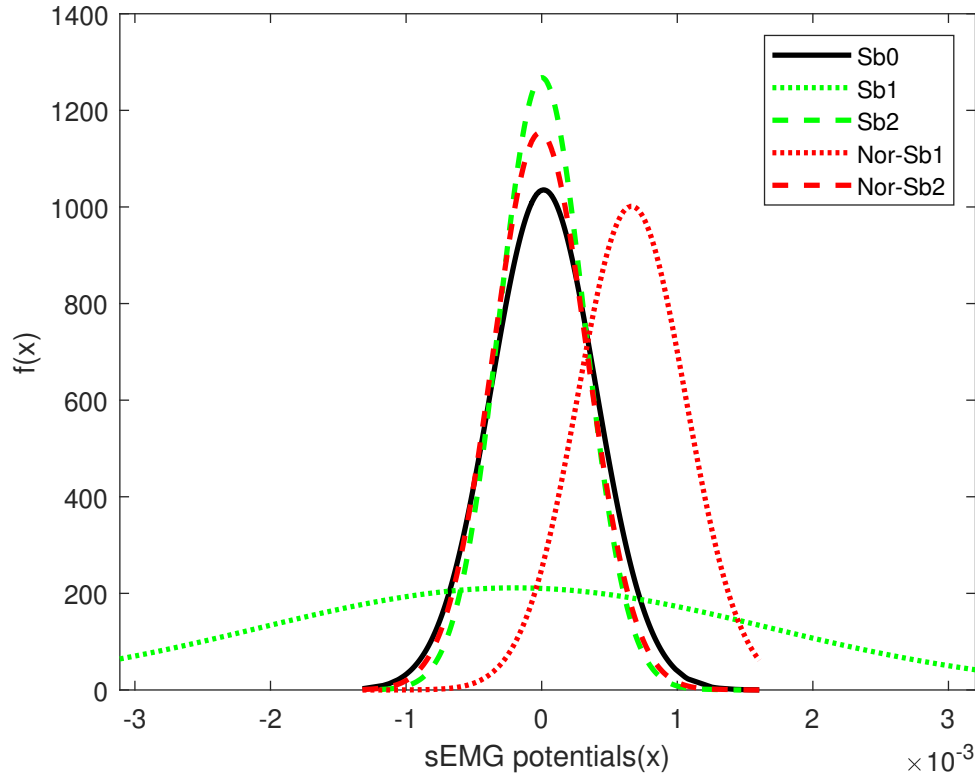


Figure 19: Probability density functions of the sEMG of the same session but different subjects

2. Taking into account a hand gesture, the maximum (max) and minimum (min) sEMG values of a channel signal collected from a user could be used to help shift and rescale the distribution of the same channel data recorded from other subjects.

To further prove our hypotheses, we employed a Gaussian Probability Density Function (PDF), shown in Equation 25, to present how dense the probability is close to a discrete sEMG magnitude. It should be noted that sEMG signals themselves may not always conform strictly to Gaussian distributions because they are complex and can be easily affected by many factors such as muscle fatigue and crosstalk. However, in practice, researchers often use Gaussian assumptions as a simplification to apply well-established statistical techniques, enabling hypothesis

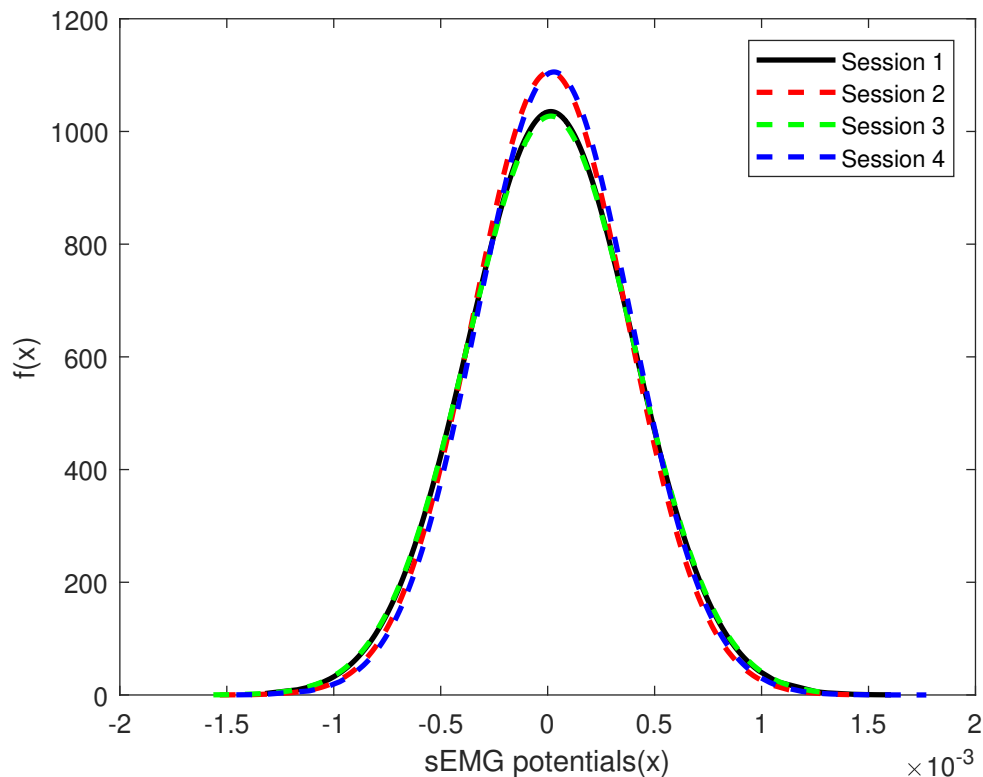


Figure 20: Probability density functions of the sEMG from the same subject but different sessions

testing, classification, and comparison of different features extracted from the signals. For example, Parker, Stuller and Scott have shown that Electromyography (EMG) signals can be modelled with a Gaussian process when they are recorded at reasonably low contraction levels.

Here we analysed the first channel sEMG signal labelled Gesture 1 (Extension) from *G.1*. As shown in Figure 20, the distributions of the first channel signal collected by the first subject (Sb0) vary only little between different sessions, while significant differences in the distributions of the same channel signal among different subjects could be found, shown by the green lines in Figure 19. It proves our first hypothesis. The red dotted and dashed plots in Figure 19 show that differences in the distributions of the same channel signal between subjects could

be reduced by applying normalisation, which proves our second hypothesis.

$$f(x) = \frac{e^{-(x-\mu)^2/(2\sigma^2)}}{\sigma\sqrt{2\pi}} \quad (25)$$

where x is a channel of sEMG onset, μ and σ are the mean and standard deviation of x .

6.4 Methods

6.4.1 Referencing normalisation

Based on the hypotheses presented above, we propose a normalisation method to reduce the domain shift problem of the inter-subject sEMG-based HGR. It is assumed that a *cycle* sEMG signal from a target subject has been collected and, therefore, can be used as reference data. The maximum and minimum sEMG potentials across all channels for each class would then be calculated from this *cycle* shown as $\max(\mathbf{X}_{i(k)}^t)$ and $\min(\mathbf{X}_{i(k)}^t)$ in Equation 26. Moreover, all *cycles* from the source domain would be normalised to reduce the domain shift. The detailed equation for the proposed normalisation approach can be seen as follows:

$$\mathbf{X}_{i(k)}^{s'} = \frac{\mathbf{X}_i^s - \min(\mathbf{X}_i^s)}{\max(\mathbf{X}_i^s) - \min(\mathbf{X}_i^s)} \times (\max(\mathbf{X}_{i(k)}^t) - \min(\mathbf{X}_{i(k)}^t)) + \min(\mathbf{X}_{i(k)}^t) \quad (26)$$

where \mathbf{X}' is a normalised sEMG onset; \mathbf{X}^s and \mathbf{X}^t represent a raw sEMG onset in the source and target domain; i and k represent the channel number and class label. The details of how to apply this normalisation approach are presented in subsection 6.4.1.

6.4.2 The model architecture

The CNN architecture implementation (in PyTorch v.1.1.0 and Python 3.7.3) for *M. Dataset* is as described in (Côté-Allard et al. 2019). It needs to be modified for *G. Dataset* since the input and output are different. The input data correspond to 16×300 (that is, the number of channels \times the number of sample points) and 8×52 while the output refers to twelve and seven hand gestures for *G. Dataset* and *M. Dataset*, respectively. The details of the modified CNN can be found in Table 13. It consists of two convolutional layers and one fully connected layer using recent techniques such as Batch Normalisation (BN) (Ioffe and Szegedy 2015), Parametric Rectified Linear Unit (PReLU) activation function (He et al. 2015b) and Adaptive Moment Estimation (ADAM) (Kingma and Ba 2015). The first and second convolutional layer has 32 and 64 filters of size 3×5 , respectively. The cross-entropy loss function is used here, and the hyperparameters: learning rate, batch size, and epoch number are selected as 0.023, 256, 100, respectively.

Table 13: Proposed CNN architecture for *G. Dataset*

Layer Order - Type	Output Shape	Parameters
L1-Conv2d	[-1, 32, 14, 296]	512
L2-BatchNorm2d	[-1, 32, 14, 296]	64
L3-PReLU	[-1, 32, 14, 296]	32
L4-Dropout2d	[-1, 32, 14, 296]	0
L5-MaxPool2d	[-1, 32, 14, 98]	0
L6-Conv2d	[-1, 64, 12, 94]	30,784
L7-BatchNorm2d	[-1, 64, 12, 94]	128
L8-PReLU	[-1, 64, 12, 94]	64
L9-Dropout2d	[-1, 64, 12, 94]	0
L10-MaxPool2d	[-1, 64, 12, 31]	0
L11-Linear	[-1, 500]	11,904,500
L12-BatchNorm1d	[-1, 500]	1,000
L13-PReLU	[-1, 500]	500
L14-Dropout	[-1, 500]	0
L15-Linear	[-1, 12]	6,012

6.4.3 System structure

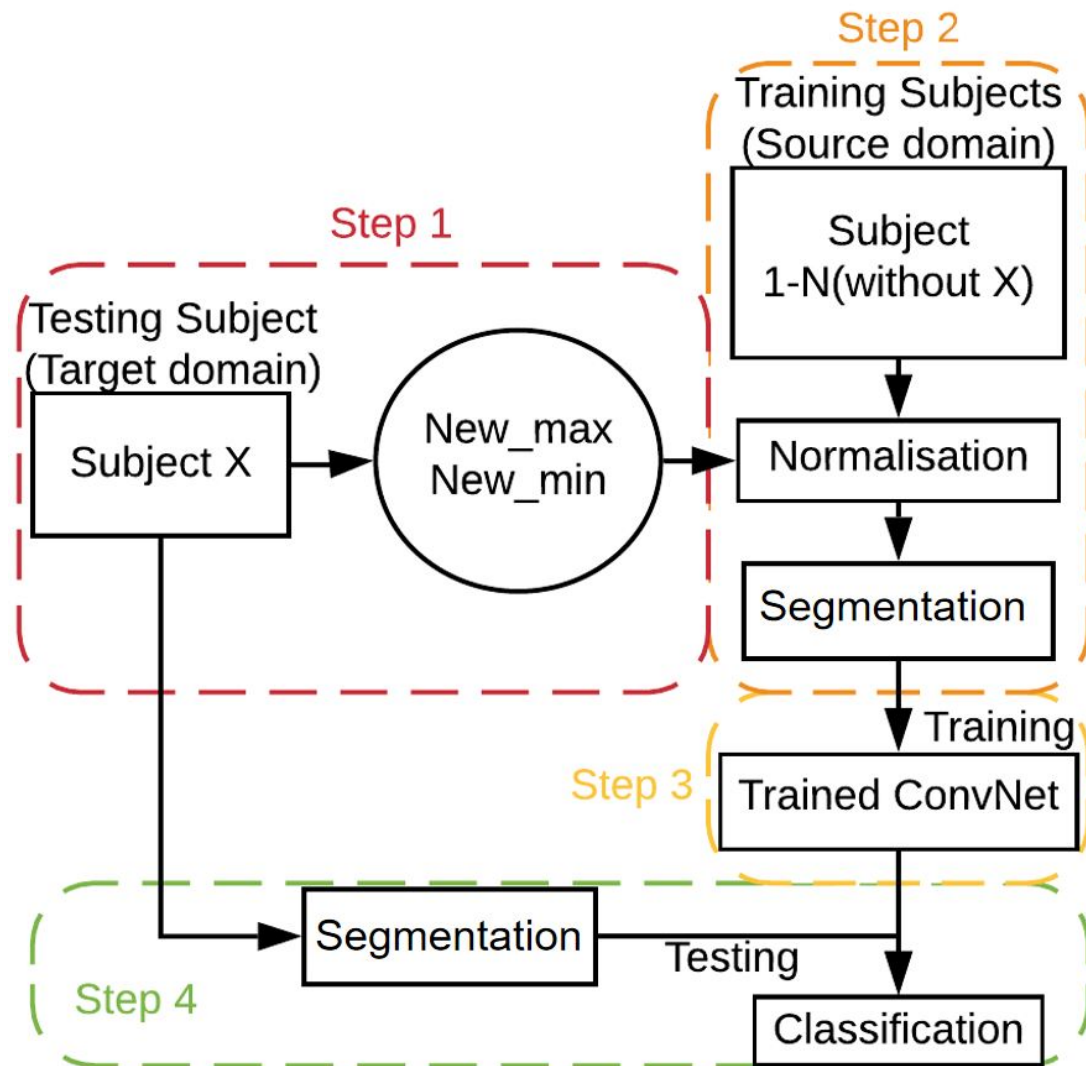


Figure 21: A high-level diagram illustrating the steps for inter-subject sEMG-based hand gesture classification

An overview working scheme for inter-subject sEMG-based HGR with a CNN using the referencing normalisation approach that we proposed could be seen in Figure 21. The first step is to calculate the maximum and minimum sEMG potentials across all channels for each class from a *cycle* collected from a testing subject. The next step is to normalise each sEMG onset in the source domain

before applying the sliding window for segmentation. After finishing the training with normalised segmented data, all other *cycles* except *cycle* 1 in the target domain are simply segmented to test the classification performance with the trained CNN.

6.5 Results

To show the robustness of the results, we used LOSOCV to evaluate the performance, and the average classification accuracy is presented without performing a majority vote. In other words, the recognition accuracy shows how many segmented test windows are correctly classified. Note that all experiments are reported as an average of 20 runs. We first investigate the performance of inter-subject sEMG-based HGR with our proposed method according to Figure 21 on *G.0*, *G.1*, and *M.0*. The results ('RNor') have been compared with using a standard normalisation ('Nor') in Table 14. Here, a standard normalisation refers to normalising each input data, that is, each segmented channel sEMG in both the source and target domains, to $[0, 1]$.

Table 14: Classification accuracy of the CNNs on three datasets with respect to the normalisation methods applied

Dataset		<i>G.0</i>	<i>G.1</i>	<i>M.0</i>
Nor	mean	33.06%	30.56%	73.76%
	std	11.03%	9.61%	9.98%
	p0	$4.08e - 09$	$7.11e - 09$	$4.12e - 12$
RNor	mean	85.09%	88.97%	94.53%
	std	6.96%	7.95%	4.79%
	p0	$3.70e - 05$	$1.13e - 05$	$3.80e - 16$
p1		$3.90e - 18$	$3.90e - 18$	$1.78e - 57$

p0: The p-value for the Shapiro-Wilk test.

p1: The p-value for the non-parametric Wilcoxon signed-rank test.

Since each result deviated from the normal distribution shown using a Shapiro-Wilk test ($n = 100, 340$ for $G.$ and $M.$ datasets) where all p values (p_0) were less than 0.05, a non-parametric Wilcoxon signed-rank test was then used to compare the classification accuracies between the normalisation methods. It can be seen that all the p-values (p_1) are less than 0.05 as in Table 14, which implies that the average recognition accuracy using our proposed normalisation was significantly higher compared to the standard normalisation across different datasets.

Additionally, we further investigated the performance of gesture classification using the recently proposed TL method (Côté-Allard et al. 2019). The first three *cycles* from training subjects were used to pre-train the CNN and then the last *cycle* was used for validation. This pre-trained model is called *Source Network*, where the parameters are frozen except for BN parameters to ensure the ability to update domain-related information from the target domain. Furthermore, a *Second Network* was trained by *cycles* from the target subject, and then merged with the *Source Network* by an element-wise summation. The different number of *cycles* (i.e., *cycle 1*; *cycle 1, 2*; *cycle 1, 2, 3*) of the target subject were used to train *Second Network* in turn to investigate how classification performance could be affected by increased training data from the target domain. To test the performance of the trained CNN using TL, the *cycle 4* of the target subject was used.

Table 15: Classification accuracy of CNNs on three datasets with raw data and transfer learning

No. of Training <i>Cycles</i>		1	2	3
$G.0$	mean	12.33%	63.59%	73.11%
	std	6.81%	24.10%	22.37%
$G.1$	mean	13.7%	44.05%	92.50%
	std	8.44%	39.58%	12.50%
$M.0$	mean	94.61%	97.41%	98.77%
	std	5.39%	3.79%	2.36%

It can be noticed from Table 15 that higher classification accuracy was achieved

as the number of training *cycles* increases. However, with the limited number of training *cycles*, the TL performs poorly in *G. Dataset*. This implies that TL requires more training *cycles* from the target domain to update domain-related information as the difficulty of the HGR task increases.

Overall, Table 14 and Table 15 show that the best performances achieved on *G.1* are higher than *G.0* regardless of the methods applied. It implies that muscle variability between subjects could be reduced if all subjects performed a hand gesture in a stable way, that is, with their elbow on instead of off the chair arm.

Furthermore, it is difficult to compare our referencing normalisation approach with the SoA TL directly due to the different number of training and testing *cycles* involved even for the same dataset. However, in terms of the inter-subject sEMG-based HGR, applying our referencing normalisation approach could achieve similar results when three *cycles* from the target subjects are used in the training process with the SoA TL.

6.6 Conclusion

This chapter presented a referencing normalisation approach to reduce domain shift when considering the inter-subject sEMG-based HGR. The classification performance using our method was investigated with a CNN on three sub datasets by LOSOCV. The subject-independent classification accuracy 85.09%, 88.97% and 94.53% was achieved in *G.0* and *G.1* when predicting 12 hand gestures not including rest and in *M.0* to classify 7 hand gestures including rest. Additionally, we applied a SoA TL method to investigate its ability to solve the domain shift problem. Our results showed that it required only one training *cycle* in *M.0* but three training *cycles* in *G.0* and *G.1* for TL to update the domain-related information so that achieving better classification accuracy than our proposed normalisation

method.

However, inter-subject sEMG-based HGR usually does not practically allow the target domain data to be involved in the training process. Compared to TL, our proposed referencing normalisation method only requires one *cycle* from a user to extract a little prior information, but it is not used for training. This is easy to implement with low computational cost. The results showed that a general high accuracy in HGR accuracy could be obtained in terms of different number of hand gestures, subjects, recording positions, and acquisition setups. Nevertheless, our method may not work well if a user is performing a hand gesture in a different way in the real world.

Chapter 7

Conclusions and Future Research

7.1 Conclusions

This thesis has a specific main focus on how to improve the robustness of Surface Electromyography (sEMG)-based Hand Gesture Recognition (HGR) by taking advantage of the most recent advanced deep learning techniques. The main research gap addressed by this study is that the tremendous success in achieving a very high sEMG-based HGR accuracy ($\geq 90\%$) reported in scientific articles produced only limited clinical or commercial impact. In light of this and the recent finding that the real-time performance of sEMG-based HGR tends to degrade significantly, these confirm the need to improve the robustness of sEMG-based HGR.

To address this problem, this thesis is concerned with the development of uncertainty-aware sEMG-based hand gesture classifiers based on the core hypothesis that the robustness of HGR can be improved by rejecting uncertain predictions. In particular, the following research questions were addressed in this thesis.

1. What does uncertainty awareness mean? Why is this considerable? And

how do we design uncertainty-aware sEMG-based hand gesture models?

2. How do we evaluate the ability of an uncertainty-aware model in a fair and comparable way? And how this can reform the current evaluation system of sEMG-based HGR?
3. How can we leverage the uncertainties inferred by uncertainty-aware models in practical use?

To address the first question, Chapter 3 presents a framework for designing an uncertainty-aware sEMG-based hand gesture classifier with design principles, theoretical details, model training approaches, hyperparameter determination process, and characteristics of inferred uncertainties. Uncertainty awareness is simply referred to the ability of a model to make inferences along with explainable uncertainties about them. This functionality has great potential to improve the robustness of the model as a general solution, rather than only to manage a particular robustness problem. Furthermore, it shows how the designed uncertainty-aware model, Evidential Convolutional Neural Network (ECNN), can be proposed by integrating a Convolutional Neural Network (CNN) with Evidential Deep Learning (EDL) without increasing the complexity of the model.

Chapter 4 mainly addresses the second question. The ability of an uncertainty-aware model is evaluated by investigating the quality of uncertainty estimates, that is, the reliability of the model based on our definition. This can be quantified by introducing the misclassification detection task. To ensure that it can be comparable in a fair way, the normalised Area Under Precision-Recall Curve (nAUPRC) is employed for reliability analysis. From the evaluation results and Accuracy-Rejection Curve (ARC) analysis shown in Section 4.8.2 and Section 4.9, it can be seen that the proposed reliability analysis has shown great potential to be a supplementary measure when evaluating sEMG-based hand gesture classifiers.

The answer to the third question can be found in Chapter 4 and Chapter 5. First, the illustration shown in Section 4.5 indicates how to investigate the rejection-capable sEMG-based finger movement recognition performance with uncertainty measures. The comparison results between CNN and ECNN briefly show the great power of the evidential uncertainties. Next, to further explore the potential of ECNN, it has been designed to solve a very challenging task, i.e., to improve the long-term robustness of hand grasp recognition. The experimental results in the challenging Non-Invasive Adaptive Hand Prosthetics (NinaPro) database 6 shown in Section 5.4.3 support our hypothesis that the robustness of sEMG-based HGR can be improved by rejecting uncertain predictions.

Finally, to address the practical concern that there is no or only limited training data available from the user, a normalisation approach has been proposed to improve the inter-subject sEMG-based HGR by reducing the subject variability of sEMG signals. Its robustness has been verified by the comparison results between this and a state-of-the-art (SoA) Transfer Learning (TL) method using Leave-One-Subject-Out Cross-Validation (LOSOCV) shown in Section 6.5.

In Chapter 6, the work presented was conducted during the early stage of my PhD research, at a time when the datasets used in the previous chapters were not yet known. As my research progressed, I gained access to the specific datasets utilised in the earlier chapters, which allowed me to perform more comprehensive analyses and draw meaningful comparisons with the subsequent findings. In specific detail, the choice of datasets in each chapter was carefully made to align with the respective research objectives and account for the varying data availability at different stages of the study. For instance, in Chapter 4, the NinaPro Database5, exercise A, was selected for reliability analysis. This particular sub dataset contains dexterous finger movements that are often challenging to distinguish from one another, making it an ideal choice for assessing the reliability of the proposed models. In Chapter 5, the NinaPro Database6 was chosen for long-term robustness

analysis due to its suitability for repeatability assessments, featuring high variability in data acquisition. Moreover, for the inter-subject recognition presented in Chapter 6, the *M. dataset* was used, featuring a relatively large number of subjects, totaling 17 individuals. Lastly, the *G. dataset*, despite comprising only 5 subjects, was selected for its unique characteristic of encompassing hand gestures performed under two different conditions. These thoughtfully chosen datasets not only enhance the comprehensiveness of our research but also allow for valuable insights and robust evaluations across different aspects of sEMG-based HGR.

The main original contributions of this thesis are summarised below.

- **Proposing a framework to design a simple but effective uncertainty-aware sEMG-based hand gesture classifier, given in Chapter 3, publications (Lin et al. 2022, 2023).**

The designed uncertainty-aware model is referred to ECNN. Its ability to make its inference along with explainable quantified multidimensional uncertainties can help us to understand the processing mechanisms of HGR to better ensure that such a reliable classifier can respond to clinical concerns. The uncertainty-aware model can provide better model interpretability while the proposed framework can improve model design efficiency in the field of sEMG-based HGR. This is achieved by helping a model achieve the ability to infer understandable multidimensional uncertainties without increasing model complexity.

- **Providing reliability analysis for sEMG-based HGR by introducing a proper definition of model reliability, as shown in Chapter 4, publication (Lin et al. 2023).**

This fills the gap of lacking consensus on the definition of model reliability in this field and offers a new dimension of evaluation, which can help model

selection together with recognition accuracy. The proposed reliability analysis has shown great potential to be a supplementary measure to evaluate sEMG-based HGR. Leveraging it can simply allow us to know which model is more reliable especially when they have similar recognition accuracy. Furthermore, the quantification of model reliability can provide an indication of rejection-capable recognition performance, which is highly dependent on the optimal rejection threshold determination and should be evaluated by more than one measure, such as the accuracy-rejection curve and false activation error. This highlights the efficiency of the proposed reliability analysis because it does not suffer from these problems.

- **Improving the long-term robustness of sEMG-based hand grasp recognition with proposed ECNN by rejecting uncertain predictions, as shown in Chapter 5, publication (Lin et al. 2023).**

Despite the fact that the potential of ECNN can be investigated by reliability analysis without optimal rejection threshold determination, one may be eager to know how the performance of HGR can be improved under rejection schemes in practical use. This study presents a practical way to determine a rejection threshold using only validation sets and to provide a comprehensive analysis of the long-term performance of rejection-capable hand grasp recognition with raw sEMG. To further explore the potential of ECNN, 3D ECNN is designed as a novel model architecture to extract both temporal and spatial features from sEMG signals. The results on a very challenging benchmark dataset, NinaPro DB6, show that both 2D and 3D ECNN outperformed the existing SoA under either non-rejection or rejection scheme. This further proves that the robustness of sEMG-based HGR can be improved by reliable models.

- **Proposing a normalisation approach to leverage large amounts of**

inter-subject data for model training so that improving recognition accuracy of sEMG-based HGR, as shown in Chapter 6, publication (Lin et al. 2020).

A novel method has been proposed to address the problem that no sufficient training data from the user is available in real-life. The inter-subject sEMG-based HGR addresses this problem by leveraging data from others but suffers from the high subject variability of sEMG signals. The results on several datasets with LOSOCV show that the domain shift can be reduced by the proposed normalisation method. The potential of our method has also been verified by presenting higher inter-subject recognition accuracy than the SoA TL, especially when only one cycle of data from the user is available.

7.2 Directions for Future research

This thesis mainly addresses a challenging research problem of how to improve the robustness of sEMG-based HGR by rejecting uncertain predictions. Despite the comprehensive results presented in Section 4.8 and Section 5.4.3 showing that the proposed uncertainty-aware model, ECNN, is more reliable and robust compared to CNN, it has much room to be improved and the current cutting-edge model will become obsolete in the future. Therefore, what we want readers to take away from this work is not just a specific model, but a theoretical framework to design reliable models and use them to tackle real-life problems, for example, to help implement restorative natural and robust upper-limb prosthetic control for amputees. Based on this, the directions for future research to explore reliable sEMG-based hand gesture classifiers are discussed from five perspectives that are relevant to model reliability: input source, model design, model training, post-processing, and applications.

7.2.1 Multimodal fusion for improving model reliability

From the perspective of model input, this thesis only considers improving the performance of sEMG-based HGR taking the input of sparse multichannel raw sEMG signals for practical concerns such as minimising computational and data acquisition cost. This can be referred to as the unimodal sEMG-based HGR. When these costs can be afforded or the improvement of model performance has a higher priority than the other factors, the multimodal fusion of sEMG and other types of data streams such as forearm kinematics (e.g., Inertial Measurement Units (IMUs)), other biomedical signals (e.g., Electroencephalogram (EEG)), and visual information (e.g., eye tracking data), can possibly be used to improve model reliability and thus produce more robust performance compared to unimodal sEMG data streams.

Recently, many studies have shown that multimodal sEMG-based HGR achieved better performance than unimodal. For example, Wei et al. shows that the proposed multi-view CNN with IMU data streams obtained an overall higher recognition accuracy in a series of benchmark datasets than only taking sEMG as input. Furthermore, Cognolato et al. presents a multimodal approach that exploits eye-hand coordination parameters with sEMG signals to improve the robustness of sEMG-based HGR. The experimental results in both transradial amputees and able-bodied subjects show that the proposed multimodal approach significantly outperforms the one relying only on the sEMG modality. It seems that multimodal sEMG-based HGR is one of the promising research directions. From the perspective of this thesis, one can investigate if model reliability can be improved by employing more modalities. If the answer is yes, it is interesting to ask the question: can we identify which modality contributes more with the aid of our proposed framework of reliability analysis? Or more importantly, how do different modalities contribute to model reliability and why? Furthermore, when introducing uncertainty-aware models for multimodal sEMG-based HGR, one can consider

using extra information from a modality to further adjust predictive uncertainties.

Rather than determining modalities, finding a way to implement multimodal fusion is also a challenging problem. As shown in Figure 22, different basic multimodal fusion strategies can be implemented and accommodated when considering combining modalities in an efficient way. Our proposed framework of reliability analysis can only provide a fair comparison between models. As such, how to find the most reliable multimodal fusion strategy remains an open question. This is important especially when more fusion strategies need to be explored as more heterogeneous data becomes available in the future.

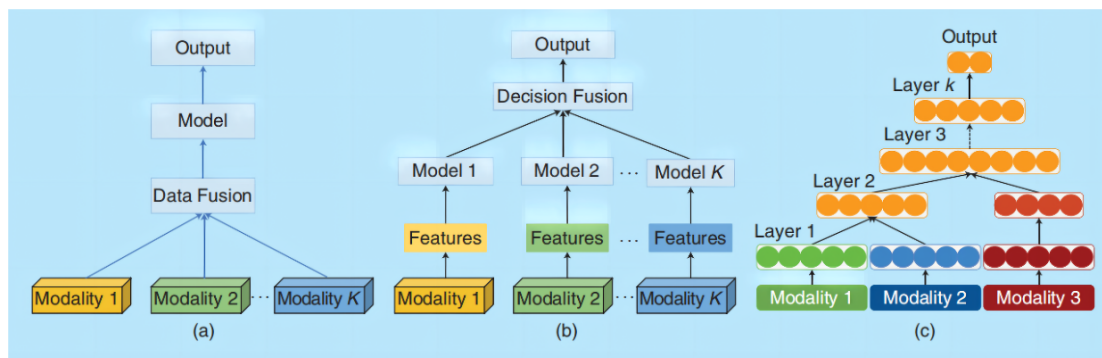


Figure 22: An illustration of different multimodal fusion strategies. (a) Early or data-level fusion, (b) late or decision-level fusion, and (c) intermediate fusion. (Rama-chandram and Taylor 2017)

7.2.2 Uncertainty-aware model design to extract advanced spatial-temporal features

Model design is always iterative and thus becomes a possible future research direction in this field. One limitation of this thesis is that we did not attempt to develop complicated sEMG-based uncertainty-aware models. For example, even though a 3D ECNN was proposed, as presented in Figure 14, to improve the long-term robustness of hand grasp recognition by extracting both spatial and temporal features, this can still be considered a simple CNN structure. It is well

known that a basic CNN can only consider local spatial features in Euclidean space. Precisely speaking, the information of sEMG electrode placement can be represented explicitly in the form of the graph network rather than a two-dimensional grid.

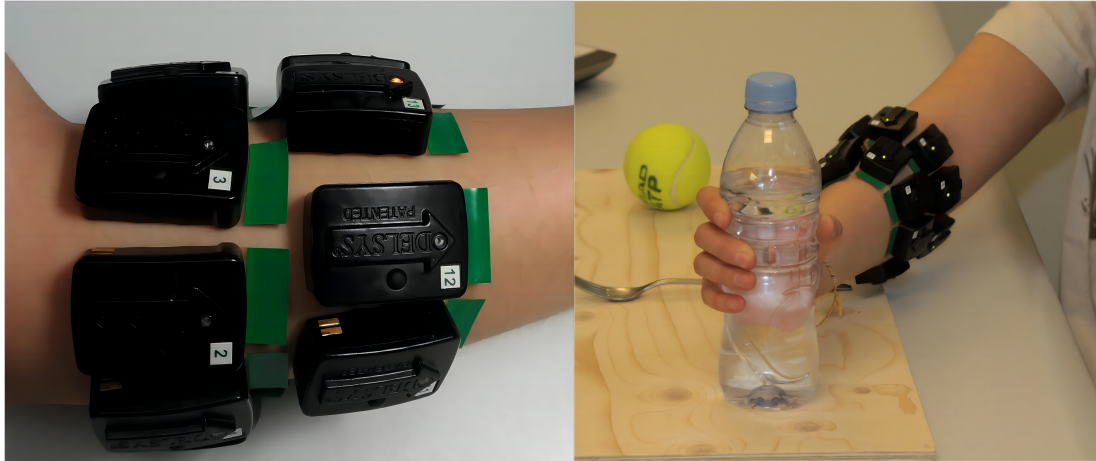


Figure 23: A collage comprised of all available pictures from (Francesca Palermo 2017) that contain the information of electrode placement for NinaPro DB6

Future research can potentially address this by exploring uncertainty-aware models based on Graph Convolutional Network (GCN), which has recently received widespread attention to handle arbitrary graph-structured data. Taking the NinaPro DB6 as an example, the detailed position of 14 electrodes can be seen in Figure 23, where the first row contains eight electrodes arranged in correspondence to the radio-humeral joint while the second row contains six electrodes placed just below the first row. To better reflect the complex topological structure of electrode placement, a graph-structured sEMG channels can be seen in Figure 24. This electrode network modelled by a general graph with a properly designed GCN may help accurately capture spatial dependence thus allowing fully utilising the spatial information during model design.

In terms of learning temporal features, a basic CNN can only capture short-term time-dependent information because of the fixed convolutional kernel size

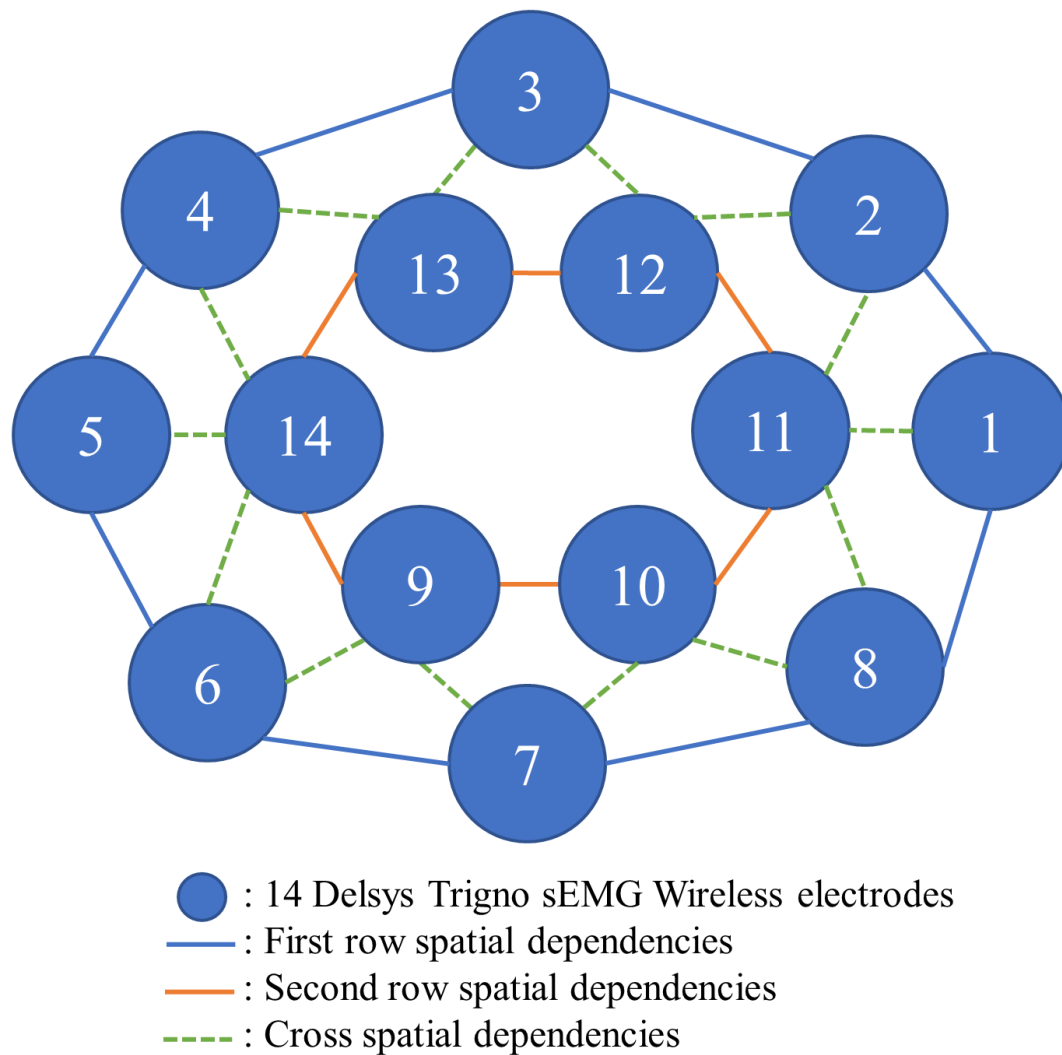


Figure 24: A graph network representing the physical spatial dependency of 14 electrodes used in NinaPro DB6

at each layer. The Temporal Convolutional Network (TCN) was carried out as a solution to address this problem by using structures such as causal dilated convolution and residual connection (Betthausen et al. 2019; Zanghieri et al. 2020a). It is interesting to design models that can extract the advanced spatial-temporal features of sEMG simultaneously, which are highly expected to enhance the overall performance of sEMG-based HGR, especially when integrating them with EDL.

7.2.3 Improve the explicit training of ECNN

It is clear that the training process of a model directly determines its performance, especially for deep learning models. As stated in Section 5.5, the general poor model reliability achieved with the uncertainty of dissonance reveals a limitation in the training of ECNN variants. To improve training for ECNN, one possible future direction is to design loss functions that can establish training strategies to penalise training samples that are misclassified differently depending on their uncertainty types. This may also be achieved by introducing AdaBoost which increases weights for misclassified samples while decreasing weights for correctly classified ones in every iteration (Wu and Nagahashi 2014). It is interesting to investigate how such weights can be adjusted based on the quantified uncertainties in different dimensions. Another direction for future research is to train a ECNN that has the ability not only to achieve high recognition accuracy in in-distribution samples but also to detect out-of-distribution samples. This can help further improve the current definition of model reliability by introducing the task of out-of-distribution detection.

7.2.4 Improve the predictive uncertainties

Future research directions in post-processing include introducing uncertainty fusion, as well as iterative uncertainty updating. Uncertainty fusion can help generate a summarised predictive uncertainty by merging uncertainties from different models with a single modality or multi-modalities (see the (b) of Figure 22). This is assumed to increase the reliability of sEMG-based HGR. However, there are many challenges that need to be addressed with care. For example, a natural question can be raised: If each model shares the same weight during the fusion process? Furthermore, how to determine the correct fusion operator, which depends on the situation, remains for further investigation for sEMG-based HGR.

Bayesian fusion is known to be one of the most commonly used postprocessing techniques for sEMG-based HGR. One interesting future research direction is to develop a Bayesian fusion post-processing approach that can integrate the multidimensional uncertainties generated by ECNN. It should be noted that the Bayesian fusion scheme can only be used when the statistical independence of the entities being combined is assumed (Kuncheva 2014). Therefore, the disjoint window should be considered instead of the sliding window with overlap when segmenting the sEMG signals.

7.2.5 Improve inter-subject recognition with uncertainty

In Chapter 6, we developed a normalisation approach to improve inter-subject sEMG-based HGR by reducing the variability of sEMG signals collected from different users with a CNN. One may ask if ECNN can be used to enhance performance. We have presented how predictive uncertainties of ECNN can be leveraged to improve rejection-capable performance of sEMG-based HGR, but did not yet explore its potential in inter-subject sEMG-based HGR. It would be interesting to train a ECNN with only limited data (e.g., one cycle) from the user and use its uncertainty-aware ability to determine if the subject variability has been reduced by our proposed normalisation method. In other words, the trained ECNN can be used to decide if samples collected from others can be used as training data for the user. If the predictive uncertainty of a normalised sample is presented as high, this suggests that this sample should not be involved in model training. Otherwise, it means that the domain shift has been reduced well and that the normalised samples (from others) can be regarded as the samples collected from the user. Under this condition, it is expected to see the inter-subject recognition performance can be improved with a large available amount of well-selected training data.

7.2.6 Real-time myoelectric control interfaces

It is hoped that the work presented in this thesis will be beneficial for designing robust and reliable myoelectric control interfaces. This is especially interesting for rehabilitation applications such as controlling assistive robots or prosthetic limbs. Note that for prosthetic limb control, our results should be confirmed by involving amputee subjects, and ideally with a human-in-the-loop experiment, in future studies. Furthermore, the performance of myoelectric control should be evaluated with real-time, closed-loop experiments when possible to pursue the ultimate goal of achieving restorative natural upper-limb function for amputees.

Bibliography

- Abdoli-Eramaki, M., Damecour, C., Christenson, J. and Stevenson, J. (2012). The effect of perspiration on the sEMG amplitude and power spectrum. *J Electromyogr Kinesiol*, 22(6), pp. 908–913.
- Abreu, J. G., Teixeira, J. M., Figueiredo, L. S. and Teichrieb, V. (2016). Evaluating sign language recognition using the Myo armband. In *SVR*, IEEE Computer Society, pp. 64–70.
- Adrian, E. D. and Bronk, D. W. (1929). The discharge of impulses in motor nerve fibres: Part ii. the frequency of discharge in reflex and voluntary contractions. *J Physiol*, 67(2), pp. i3–151.
- Ajiboye, A. and Weir, R. (2005). A heuristic fuzzy logic approach to EMG pattern recognition for multifunctional prosthesis control. *IEEE Trans Neural Sys & Rehab Eng*, 13(3), pp. 280–291.
- Akhmadeev, K. (2019). *Probabilistic models based on EMG decomposition for prosthetic control*. Ph.D. thesis, University Bretagne Loire.
- Akiba, T., Sano, S., Yanase, T., Ohta, T. and Koyama, M. (2019). Optuna: A next-generation hyperparameter optimization framework. In *KDD*, ACM, pp. 2623–2631.
- Al-Timemy, A. H., Khushaba, R. N., Bugmann, G. and Escudero, J. (2016).

- Improving the performance against force variation of EMG controlled multi-functional upper-limb prostheses for transradial amputees. *IEEE Trans Neural Syst Rehabil Eng*, 24(6), pp. 650–661.
- Ameri, A., Akhaee, M. A., Scheme, E. and Englehart, K. (2019). Regression convolutional neural network for improved simultaneous EMG control. *J Neural Eng*, 16(3), p. 036015.
- Amsüss, S. et al. (2014). Self-correcting pattern recognition system of surface EMG signals for upper limb prosthesis control. *IEEE Trans Biomed Eng*, 61(4), pp. 1167–1176.
- Ashukha, A., Lyzhov, A., Molchanov, D. and Vetrov, D. P. (2020). Pitfalls of in-domain uncertainty estimation and ensembling in deep learning. In *ICLR*, OpenReview.net, p. 4.
- Atzori, M., Cognolato, M. and Müller, H. (2016). Deep learning with convolutional neural networks applied to electromyography data: A resource for the classification of movements for prosthetic hands. *Front Neurobot*, 10, p. 9.
- Atzori, M. and Müller, H. (2015). Control capabilities of myoelectric robotic prostheses by hand amputees: A scientific research and market overview. *Front Syst Neurosci*, 9, p. 162.
- Atzori, M. and Müller, H. (2015). The Ninapro database: A resource for sEMG naturally controlled robotic hand prosthetics. In *EMBC*, IEEE, pp. 7151–7154.
- Atzori, M. et al. (2014). Electromyography data for non-invasive naturally-controlled robotic hand prostheses. *Sci Data*, 1(1), pp. 1–13.
- Bai, S., Kolter, J. Z. and Koltun, V. (2018). An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *CoRR*, abs/1803.01271, pp. 1–14.

- Bao, T., Zaidi, S. A. R., Xie, S. Q., Yang, P. and Zhang, Z.-Q. (2022). CNN confidence estimation for rejection-based hand gesture classification in myoelectric control. *IEEE Trans Hum Mach Syst*, 52(1), pp. 99–109.
- Basmajian, J. V. and Stecko, G. (1962). A new bipolar electrode for electromyography. *J Appl Physiol*, 17(5), pp. 849–849.
- Beaulieu, R. J. et al. (2017). Multi-position training improves robustness of pattern recognition and reduces limb-position effect in prosthetic control. *J Prosthet Orthot*, 29(2), pp. 54–62.
- Beniczky, S., Conradsen, I. and Wolf, P. (2018). Detection of convulsive seizures using surface electromyography. *Epilepsia*, 59, pp. 23–29.
- Bergstra, J., Yamins, D. and Cox, D. D. (2013). Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures. In *ICML, JMLR Workshop and Conference Proceedings*, vol. 28, JMLR.org, pp. 115–123.
- Bergstra, J., Bardenet, R., Bengio, Y. and Kégl, B. (2011). Algorithms for hyperparameter optimization. In *NIPS*, pp. 2546–2554.
- Betthausen, J. L. et al. (2018). Limb position tolerant pattern recognition for myoelectric prosthesis control with adaptive sparse representations from extreme learning. *IEEE Trans Biomed Eng*, 65(4), pp. 770–778.
- Betthausen, J. L., Krall, J. T., Kaliki, R. R., Fifer, M. S. and Thakor, N. V. (2019). Stable electromyographic sequence prediction during movement transitions using temporal convolutional networks. In *2019 9th International IEEE/EMBS Conference on Neural Engineering (NER)*, pp. 1046–1049.
- Betthausen, J. L. et al. (2020). Stable responsive EMG sequence prediction and

- adaptive reinforcement with temporal convolutional networks. *IEEE Trans Biomed Eng*, 67(6), pp. 1707–1717.
- Blanc, Y. and Dimanico, U. (2010). History of the study of skeletal muscle function with emphasis on kinesiological electromyography. *Open Rehabil J*, 3, pp. 84–93.
- Blundell, C., Cornebise, J., Kavukcuoglu, K. and Wierstra, D. (2015). Weight uncertainty in neural network. In *ICML, JMLR Workshop and Conference Proceedings*, vol. 37, JMLR.org, pp. 1613–1622.
- Boostani, R. and Moradi, M. H. (2003). Evaluation of the forearm EMG signal features for the control of a prosthetic hand. *Physiol Meas*, 24, pp. 309–319.
- Boyd, K., Eng, K. H. and Page, D. (2013). Area under the precision-recall curve: Point estimates and confidence intervals. In *ECML/PKDD (3), Lecture Notes in Computer Science*, vol. 8190, Springer, pp. 451–466.
- Boyd, K., Davis, J., Page, D. and Costa, V. S. (2012). Unachievable region in precision-recall space and its effect on empirical evaluation. In *ICML*, Omnipress, pp. 639–646.
- Burden, A. (2010). How should we normalize electromyograms obtained from healthy participants? What we have learned from over 25years of research. *J Electromyogr and Kinesiol*, 20(6), pp. 1023–1035.
- Campbell, E., Phinyomark, A. and Scheme, E. (2019). Linear discriminant analysis with Bayesian risk parameters for myoelectric control. In *2019 IEEE GlobalSIP*, IEEE, pp. 1–5.
- Cene, V. H., Tosin, M. C., Machado, J. and Balbinot, A. (2019). Open database for accurate upper-limb intent detection using electromyography and reliable extreme learning machines. *Sensors*, 19(8), p. 1864.

- Chakrabarti, B. et al. (2015). Overview on literature survey towards EMG interpretations technique in addition to several interdisciplinary work related to EMG interpretations. *Int J Sci Eng Res*, 6(5), pp. 419–424.
- Chan, A. and Englehart, K. (2005). Continuous myoelectric control for powered prostheses using hidden markov models. *IEEE Trans Biomed Eng*, 52(1), pp. 121–124.
- Chan, F., Yang, Y.-S., Lam, F., Zhang, Y.-T. and Parker, P. (2000). Fuzzy EMG classification for prosthesis control. *IEEE Trans Rehabil Eng*, 8(3), pp. 305–311.
- Chang, J., Phinyomark, A. and Scheme, E. (2020). Assessment of EMG benchmark data for gesture recognition using the NinaPro database. In *EMBC*, IEEE, pp. 3339–3342.
- Chen, X. et al. (2007). Multiple hand gesture recognition based on surface EMG signal. In *2007 1st Int Conf Bioinform Biomed Eng*, IEEE, pp. 506–509.
- Chen, X., Guo, H., Wang, G. and Zhang, L. (2017). Motion feature augmented recurrent neural network for skeleton-based dynamic hand gesture recognition. In *IEEE ICIP*, pp. 2881–2885.
- Cheng, J., Chen, X., Liu, A. and Peng, H. (2015). A novel phonology- and radical-coded Chinese sign language recognition framework using accelerometer and surface electromyography sensors. *Sensors*, 15(9), pp. 23303–23324.
- Chowdhury, A., Ramadas, R. and Karmakar, S. (2013). Muscle computer interface: A review. In *Lecture Notes in Mechanical Engineering*, Springer India, pp. 411–421.
- Cognolato, M. et al. (2020). Gaze, visual, myoelectric, and inertial data of grasps for intelligent prosthetics. *Sci Data*, 7(1), 43.

- Cognolato, M., Atzori, M., Gassert, R. and Müller, H. (2022). Improving robotic hand prosthesis control with eye tracking and computer vision: A multimodal approach based on the visuomotor behavior of grasping. *Front Artif Intell*, 4, 744476.
- Côté-Allard, U. et al. (2019). Deep learning for electromyographic hand gesture signal classification using transfer learning. *IEEE Trans Neural Sys & Rehab Eng*, 27(4), pp. 760–771.
- Côté-Allard, U. et al. (2020a). Interpreting deep learning features for myoelectric control: A comparison with handcrafted features. *Front Bioeng Biotechnol*, 8, p. 158.
- Côté-Allard, U. et al. (2020b). Unsupervised domain adversarial self-calibration for electromyography-based gesture recognition. *IEEE Access*, 8, pp. 177941–177955.
- Côté-Allard, U. et al. (2021). A transferable adaptive domain adversarial neural network for virtual reality augmented EMG-based gesture recognition. *IEEE Trans Neural Sys & Rehab Eng*, 29, pp. 546–555.
- Dempster, A. P. (2008). A generalization of Bayesian inference. In *Classic Works of the Dempster-Shafer Theory of Belief Functions, Studies in Fuzziness and Soft Computing*, vol. 219, Springer, pp. 73–104.
- Drost, G., Stegeman, D. F., van Engelen, B. G. and Zwarts, M. J. (2006). Clinical applications of high-density surface EMG: A systematic review. *J Electromyogr Kinesiol*, 16(6), pp. 586–602.
- Du, Y., Jin, W., Wei, W., Hu, Y. and Geng, W. (2017). Surface EMG-based inter-session gesture recognition enhanced by deep domain adaptation. *Sensors*, 17(3), p. 458.

- Englehart, K. and Hudgins, B. (2003). A robust, real-time control scheme for multifunction myoelectric control. *IEEE Trans Biomed Eng*, 50(7), pp. 848–854.
- Englehart, K. B., Hudgins, B. and Parker, P. A. (2001). A wavelet-based continuous classification scheme for multifunction myoelectric control. *IEEE Trans Biomed Eng*, 48(3), pp. 302–311.
- Farina, D. et al. (2014). The extraction of neural information from the surface EMG for the control of upper-limb prostheses: emerging avenues and challenges. *IEEE Trans Neural Sys & Rehab Eng*, 22(4), pp. 797–809.
- Fatimah, B., Singh, P., Singhal, A. and Pachori, R. B. (2021). Hand movement recognition from sEMG signals using fourier decomposition method. *Biocybern Biomed Eng*, 41, pp. 690–703.
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognit Lett*, 27(8), pp. 861–874.
- Finley, F. and Wirta, R. (1967). Myocoder studies of multiple myopotential response. *Arch Phys Med Rehabil*, 48(11), pp. 598–601.
- Fitts, P. M. (1992). The information capacity of the human motor system in controlling the amplitude of movement. *J Exp Psychol Gen*, 121(3), pp. 262–269.
- Forman, G. and Scholz, M. (2010). Apples-to-apples in cross-validation studies: pitfalls in classifier performance measurement. *SIGKDD Explor*, 12(1), pp. 49–57.
- Fougner, A., Scheme, E., Chan, A. D. C., Englehart, K. and Stavdahl, Ø. (2011). Resolving the limb position effect in myoelectric pattern recognition. *IEEE Trans Neural Syst Rehabil Eng*, 19(6), pp. 644–651.

- Fougner, A. L., Stavadahl, Ø., Kyberd, P. J., Losier, Y. G. and Parker, P. A. (2012). Control of upper limb prostheses: Terminology and proportional myoelectric control—A review. *IEEE Trans Neural Syst Rehabil Eng*, 20, pp. 663–677.
- Francesca Palermo, M. A., Barbara Caputo (2017). *Repeatability analysis of hand movement recognition for control of robotic prosthesis based on sEMG data*. Master's thesis, The Sapienza University of Rome.
- Fukuda, O., Tsuji, T., Kaneko, M. and Otsuka, A. (2003). A human-assisting manipulator teleoperated by EMG signals and arm motions. *IEEE Trans Robot Autom*, 19(2), pp. 210–222.
- Gasser, H. S. and Erlanger, J. (1922). A study of the action currents of nerve with the cathode ray oscillograph. *American Journal of Physiology-Legacy Content*, 62(3), pp. 496–524.
- Geng, W. et al. (2016). Gesture recognition by instantaneous surface EMG images. *Sci Rep*, 6(1), 36571.
- Geng, Y., Zhou, P. and Li, G. (2012). Toward attenuating the impact of arm positions on electromyography pattern-recognition based motion classification in transradial amputees. *J Neuroeng Rehabil*, 9, 74.
- Geng, Y., Zhang, F., Yang, L., Zhang, Y. and Li, G. (2012). Reduction of the effect of arm position variation on real-time performance of motion classification. In *2012 Annu Int Conf IEEE EMBC*, IEEE, pp. 2772–2775.
- Geng, Y., Samuel, O. W., Wei, Y. and Li, G. (2017). Improving the robustness of real-time myoelectric pattern recognition against arm position changes in transradial amputees. *Biomed Res Int*, 2017, 5090454.
- Gerilovsky, L., Tsvetinov, P. and Trenkova, G. (1986). H-reflex potentials shape

- and amplitude changes at different length of relaxed soleus muscle. *Electromyogr Clin Neurophysiol*, 26(8), pp. 641–653.
- Ghassemi, M. et al. (2019). Development of an EMG-controlled serious game for rehabilitation. *IEEE Trans Neural Sys & Rehab Eng*, 27(2), pp. 283–292.
- Golovin, D. et al. (2017). Google vizier: A service for black-box optimization. In *KDD*, ACM, pp. 1487–1495.
- Gu, Y., Yang, D., Huang, Q., Yang, W. and Liu, H. (2018). Robust EMG pattern recognition in the presence of confounding factors: features, classifiers and adaptive learning. *Expert Syst Appl*, 96, pp. 208–217.
- Hakonen, M., Piitulainen, H. and Visala, A. (2015). Current state of digital signal processing in myoelectric interfaces and related applications. *Biomed Signal Process Control*, 18, pp. 334–359.
- Halaki, M. and Gi, K. (2012). Normalization of EMG signals: To normalize or not to normalize and what to normalize to? In G. R. Naik, ed., *Computational Intelligence in Electromyography Analysis*, InTechOpen, chap. 7, pp. 175–194.
- Hargrove, L. J., Englehart, K. B. and Hudgins, B. (2006). The effect of electrode displacements on pattern recognition based myoelectric control. In *EMBC*, IEEE, pp. 2203–2206.
- Hargrove, L. J., Englehart, K. B. and Hudgins, B. (2008). A training strategy to reduce classification degradation due to electrode displacements in pattern recognition based myoelectric control. *Biomed Signal Process Control*, 3(2), pp. 175–180.

- Hargrove, L. J., Scheme, E. J., Englehart, K. B. and Hudgins, B. S. (2010). Multiple binary classifications via linear discriminant analysis for improved controllability of a powered prosthesis. *IEEE Trans Neural Sys & Rehab Eng*, 18(1), pp. 49–57.
- Hartwell, A., Kadiramanathan, V. and Anderson, S. R. (2018). Compact deep neural networks for computationally efficient gesture classification from electromyography signals. In *BioRob*, IEEE, pp. 891–896.
- Hashimoto, S. et al. (1994). Waveform changes of compound muscle action potential (CMAP) with muscle length. *J Neurol Sci*, 124(1), pp. 21–24.
- He, J. et al. (2015a). User adaptation in long-term, open-loop myoelectric training: implications for EMG pattern recognition in prosthesis control. *J Neural Eng*, 12(4), 046005.
- He, K., Zhang, X., Ren, S. and Sun, J. (2015b). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *ICCV*, IEEE Computer Society, pp. 1026–1034.
- He, Y., Fukuda, O., Bu, N., Okumura, H. and Yamaguchi, N. (2018). Surface EMG pattern recognition using long short-term memory combined with multilayer perceptron. In *2018 40th Annu Int Conf IEEE Eng Med Biol Soc*, IEEE, pp. 5636–5639.
- Herberts, P., Almström, C., Kadefors, R. and Lawrence, P. D. (1973). Hand prosthesis control via myoelectric patterns. *Acta Orthop Scand*, 44(4-5), pp. 389–409.
- Hintze, J. L. and Nelson, R. D. (1998). Violin plots: A box plot-density trace synergism. *The American Statistician*, 52(2), pp. 181–184.

- Hu, Y. et al. (2019). sEMG-based gesture recognition with embedded virtual hand poses and adversarial learning. *IEEE Access*, 7, pp. 104108–104120.
- Huang, Q. et al. (2017). A novel unsupervised adaptive learning method for long-term electromyography (EMG) pattern recognition. *Sensors*, 17(6), 1370.
- Huang, Y., Englehart, K., Hudgins, B. and Chan, A. (2005). A gaussian mixture model based classification scheme for myoelectric control of powered upper limb prostheses. *IEEE Trans Biomed Eng*, 52(11), pp. 1801–1811.
- Hudgins, B., Parker, P. and Scott, R. N. (1993). A new strategy for multifunction myoelectric control. *IEEE Trans Biomed Eng*, 40(1), pp. 82–94.
- Hutter, F., Hoos, H. H. and Leyton-Brown, K. (2011). Sequential model-based optimization for general algorithm configuration. In *LION, Lecture Notes in Computer Science*, vol. 6683, Springer, pp. 507–523.
- Hwang, H. J., Hahne, J. M. and Müller, K. R. (2017). Real-time robustness evaluation of regression based myoelectric control against arm position change and donning/doffing. *PLoS One*, 12(11), e0186318.
- Ioffe, S. and Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *ICML, JMLR Workshop and Conference Proceedings*, vol. 37, JMLR.org, pp. 448–456.
- Jaramillo-Yáñez, A., Benalcázar, M. E. and Mena-Maldonado, E. (2020). Real-time hand gesture recognition using surface electromyography and machine learning: A systematic literature review. *Sensors*, 20(9), 2467.
- Jensen, C., Vasseljen, O. and Westgaard, R. H. (1993). The influence of electrode position on bipolar surface electromyogram recordings of the upper trapezius muscle. *Eur J Appl Physiol Occup Physiol*, 67(3), pp. 266–273.

- Jiang, N., Muceli, S., Graimann, B. and Farina, D. (2012a). Effect of arm position on the prediction of kinematics from EMG in amputees. *Med Biol Eng Comput*, 51(1-2), pp. 143–151.
- Jiang, N., Dosen, S., Muller, K.-R. and Farina, D. (2012b). Myoelectric control of artificial limbs—is there a need to change focus? [in the spotlight]. *IEEE Signal Process Mag*, 29(5), pp. 152–150.
- Jøsang, A. (2016). *Subjective Logic - A Formalism for Reasoning Under Uncertainty*. Artificial Intelligence: Foundations, Theory, and Algorithms, Springer.
- Jøsang, A., Cho, J. and Chen, F. (2018). Uncertainty characteristics of subjective opinions. In *FUSION*, IEEE, pp. 1998–2005.
- Kaczmarek, P., Mańkowski, T. and Tomczyński, J. (2019). putEMG—A surface electromyography hand gesture recognition dataset. *Sensors*, 19(16), 3548.
- Kamen, G. (2013). Electromyographic kinesiology. In L. D. Robertson, A. N. Tocco, K. Maurer, S. Huls, C. A. Gentis, E. Evans and J. Anderson, eds., *Research Methods in Biomechanics*, Human Kinetics, chap. 8, pp. 179–201.
- Kaplan, L. M., Cerutti, F., Sensoy, M., Preece, A. D. and Sullivan, P. (2018). Uncertainty aware AI ML: why and how. *CoRR*, abs/1809.07882.
- Karlik, B., Tokhi, M. and Alci, M. (2003). A fuzzy clustering neural network architecture for multifunction upper-limb prosthesis. *IEEE Trans Biomed Eng*, 50(11), pp. 1255–1261.
- Karlsson, S., Yu, J. and Akay, M. (2000). Time-frequency analysis of myoelectric signals during dynamic contractions: A comparative study. *IEEE Trans Biomed Eng*, 47(2), pp. 228–238.

- Kaufmann, P., Englehart, K. B. and Platzner, M. (2010). Fluctuating EMG signals: Investigating long-term effects of pattern matching algorithms. *Conf Proc IEEE Eng Med Biol Soc*, pp. 6357–6360.
- Kerber, F., Puhl, M. and Krüger, A. (2017). User-independent real-time hand gesture recognition based on surface electromyography. In *19th MobileHCI*, ACM, pp. 1–7.
- Khokhar, Z. O., Xiao, Z. G. and Menon, C. (2010). Surface EMG pattern recognition for real-time control of a wrist exoskeleton. *Biomed Eng Online*, 9(41), pp. 1–17.
- Khushaba, R. N. (2014). Correlation analysis of electromyogram signals for multiuser myoelectric interfaces. *IEEE Trans Neural Syst Rehabil Eng*, 22(4), pp. 745–755.
- Khushaba, R. N., Kodagoda, S., Takruri, M. and Dissanayake, G. (2012). Toward improved control of prosthetic fingers using surface electromyogram (EMG) signals. *Expert Syst Appl*, 39(12), pp. 10731–10738.
- Khushaba, R. N., Takruri, M., Miró, J. V. and Kodagoda, S. (2014). Towards limb position invariant myoelectric pattern recognition using time-dependent spectral features. *Neural Netw*, 55, pp. 42–58.
- Kingma, D. P. and Ba, J. (2015). ADAM: A method for stochastic optimization. In *ICLR (Poster)*, pp. 1–15.
- Kleissen, R., Buurke, J., Harlaar, J. and Zilvold, G. (1998). Electromyography in the biomechanical analysis of human movement and its clinical application. *Gait Posture*, 8(2), pp. 143–158.
- Kopetzki, A., Charpentier, B., Zügner, D., Giri, S. and Günemann, S. (2021). Evaluating robustness of predictive uncertainty estimation: Are Dirichlet-based

- models reliable? In *ICML, Proceedings of Machine Learning Research*, vol. 139, PMLR, pp. 5707–5718.
- Krasoulis, A., Vijayakumar, S. and Nazarpour, K. (2020). Multi-grip classification-based prosthesis control with two EMG-IMU sensors. *IEEE Trans Neural Syst Rehabil Eng*, 28(2), pp. 508–518.
- Krasoulis, A., Kyranou, I., Erden, M. S., Nazarpour, K. and Vijayakumar, S. (2017). Improved prosthetic hand control with concurrent use of myoelectric and inertial measurements. *J Neuroeng Rehabil*, 14(1), 71.
- Krstajic, D., Buturovic, L. J., Leahy, D. E. and Thomas, S. (2014). Cross-validation pitfalls when selecting and assessing regression and classification models. *J Cheminformatics*, 6(1), 10.
- Kuncheva, L. I. (2014). *Combining pattern classifiers: methods and algorithms*. John Wiley & Sons.
- Kurakin, A., Goodfellow, I. J. and Bengio, S. (2016). Adversarial machine learning at scale. *CoRR*, abs/1611.01236, 1611.01236.
- Kutafina, E., Laukamp, D. and Jonas, S. M. (2015). Wearable sensors in medical education: Supporting hand hygiene training with a forearm EMG. In *pHealth 2015, Studies in Health Technology and Informatics*, vol. 211, pp. 286–291.
- Kuzborskij, I., Gijsberts, A. and Caputo, B. (2012). On the challenge of classifying 52 hand movements from surface electromyography. In *2012 Annu Int Conf IEEE Eng Med Biol Soc*, IEEE, pp. 4931–4937.
- Lakshminarayanan, B., Pritzel, A. and Blundell, C. (2017). Simple and scalable predictive uncertainty estimation using deep ensembles. In *NIPS*, pp. 6402–6413.

- Lehman, G. J. and McGill, S. M. (1999). The importance of normalization in the interpretation of surface electromyography: A proof of principle. *J Manipulative Physiol Ther*, 22(7), pp. 444–446.
- Leonardis, D. et al. (2015). An EMG-controlled robotic hand exoskeleton for bilateral rehabilitation. *IEEE Trans Haptics*, 8(2), pp. 140–151.
- Li, L., Looney, D., Park, C., Rehman, N. U. and Mandic, D. P. (2011). Power independent EMG based gesture recognition for robotics. In *2011 Annu Int Conf IEEE Eng Med Biol Soc*, IEEE, pp. 793–796.
- Li, W., Shi, P. and Yu, H. (2021). Gesture recognition using surface electromyography and deep learning for prostheses hand: State-of-the-art, challenges, and future. *Front Neurosci*, 15, 621885.
- Li, Y., Zhang, Q., Zeng, N., Chen, J. and Zhang, Q. (2019). Discrete hand motion intention decoding based on transient myoelectric signals. *IEEE Access*, 7, pp. 81630–81639.
- Lin, Y., De Wilde, P., Palaniappan, R. and Li, L. (2018). Muscle connectivity analysis for hand gesture recognition via sEMG. In *2018 Signal Inf Process Assoc Annu Summit Conf APSIPA Asia Pac. (APSIPA ASC)*, IEEE, pp. 848–852.
- Lin, Y., Palaniappan, R., De Wilde, P. and Li, L. (2020). A normalisation approach improves the performance of inter-subject sEMG-based hand gesture recognition with a ConvNet. In *EMBC*, IEEE, pp. 649–652.
- Lin, Y., Palaniappan, R., De Wilde, P. and Li, L. (2022). Reliability analysis for finger movement recognition with raw electromyographic signal by evidential convolutional networks. *IEEE Trans Neural Sys & Rehab Eng*, 30, pp. 96–107.

- Lin, Y., Palaniappan, R., De Wilde, P. and Li, L. (2023). Robust long-term hand grasp recognition with raw electromyographic signals using multidimensional uncertainty-aware models. *IEEE Trans Neural Sys & Rehab Eng*, 31, pp. 962–971.
- Lowery, M. (2009). Electromyography. In M. D. Binder, N. Hirokawa and U. Windhorst, eds., *Encyclopedia of Neuroscience*, NameSpringer, pp. 1074–1077.
- Lu, Z., Chen, X., Zhang, X., Tong, K.-Y. and Zhou, P. (2017). Real-time control of an exoskeleton hand robot with myoelectric pattern recognition. *Int J Neural Syst*, 27(05), 1750009.
- Luca, C. J. D. (2002). Surface electromyography: Detection and recording. Tech. rep., DelSys Incorporated.
- Luca, C. J. D., Adam, A., Wotiz, R., Gilmore, L. D. and Nawab, S. H. (2006). Decomposition of surface EMG signals. *J Neurophysiol*, 96(3), pp. 1646–1657.
- Marano, G. et al. (2021). Questioning domain adaptation in myoelectric hand prostheses control: An inter- and intra-subject study. *Sensors*, 21(22), 7500.
- Matsubara, T., Hyon, S.-H. and Morimoto, J. (2011). Learning and adaptation of a stylistic myoelectric interface: EMG-based robotic control with individual user differences. *IEEE ROBIO 2011*, pp. 390–395.
- Mezzarane, R. A., Elias, L. A., Magalhães, F. H., Chaud, V. M. and Kohn, A. F. (2013). Experimental and simulated EMG responses in the study of the human spinal cord. In *Electrodiagnosis in New Frontiers of Clinical Research*, InTech, pp. 57–87.
- Mohebbian, M. R. et al. (2021). A comprehensive review of myoelectric prosthesis control. <http://arxiv.org/abs/2112.13192>.

- Moin, A. et al. (2018). An EMG gesture recognition system with flexible high-density sensors and brain-inspired high-dimensional classifier. In *2018 IEEE ISCAS*, pp. 1–5.
- Montazerin, M., Zabihi, S., Rahimian, E., Mohammadi, A. and Naderkhani, F. (2022). ViT-HGR: Vision transformer-based hand gesture recognition from high density surface EMG signals. <http://arxiv.org/abs/2201.10060v1>.
- Moqadam, S. B., Elahi, S. M., Mo, A. and Zhang, W. (2018). Hybrid control combined with a voluntary biosignal to control a prosthetic hand. *Robotics Biomim*, 5(1), 4.
- Morais, G. D., Neves, L. C., Masiero, A. A. and Castro, M. (2016). Application of Myo armband system to control a robot interface. *BIOSIGNALS*, 4, pp. 227–231.
- Nadeem, M. S. A., Zucker, J. and Hanczar, B. (2010). Accuracy-rejection curves (ARCs) for comparing classification methods with a reject option. In *MLSB, JMLR Proceedings*, vol. 8, PMLR, pp. 65–81.
- Nasri, N., Orts-Escolano, S., Gomez-Donoso, F. and Cazorla, M. (2019). Inferring static hand poses from a low-cost non-intrusive sEMG sensor. *Sensors (Basel)*, 19, 371.
- Neuman, M. R. (2010). Biopotential electrodes. In J. Webster, ed., *Medical Instrumentation Application and Design*, John Wiley & Sons, chap. 5, 4th edn., pp. 189–235.
- Ng, K. W., Tian, G.-L. and Tang, M.-L. (2011). *Dirichlet and related distributions: Theory, methods and applications*. Wiley Series in Probability and Statistics, John Wiley & Sons.

- Nordander, C. et al. (2003). Influence of the subcutaneous fat layer, as measured by ultrasound, skinfold calipers and BMI, on the EMG amplitude. *Eur J Appl Physiol*, 89(6), pp. 514–519.
- Oskoei, M. A. and Hu, H. (2007). Myoelectric control systems—A survey. *Biomed Signal Process Control*, 2(4), pp. 275–294.
- Paiss, O. and Inbar, G. F. (1987). Autoregressive modeling of surface EMG and its spectrum with application to fatigue. *IEEE Trans Biomed Eng*, BME-34(10), pp. 761–770.
- Palermo, F. et al. (2017). Repeatability of grasp recognition for robotic hand prosthesis control based on sEMG data. In *ICORR*, IEEE, pp. 1154–1159.
- Papagiannis, G. I. et al. (2019). Methodology of surface electromyography in gait analysis: review of the literature. *J Med Eng Technol*, 43(1), pp. 59–65.
- Parker, P., Stuller, J. and Scott, R. (1977). Signal processing for the multistate myoelectric channel. *Proc IEEE*, 65(5), pp. 662–674.
- Phinyomark, A., Phukpattaranont, P. and Limsakul, C. (2012). Feature reduction and selection for EMG signal classification. *Expert Syst Appl*, 39(8), pp. 7420–7431.
- Phinyomark, A. and Scheme, E. (2018). EMG pattern recognition in the era of big data and deep learning. *Big Data Cogn Comput*, 2(3), 21.
- Phinyomark, A. et al. (2013). EMG feature evaluation for improving myoelectric pattern recognition robustness. *Expert Syst Appl*, 40(12), pp. 4832–4840.
- Pizzolato, S. et al. (2017). Comparison of six electromyography acquisition setups on hand movement classification tasks. *PLoS One*, 12(10), e0186132.

- Prahn, C. et al. (2019). Counteracting electrode shifts in upper-limb prosthesis control via transfer learning. *IEEE Trans Neural Sys & Rehab Eng*, 27(5), pp. 956–962.
- Rahimian, E. et al. (2021). FS-HGR: Few-shot learning for hand gesture recognition via electromyography. *IEEE Trans Neural Syst Rehabil Eng*, 29, pp. 1004–1015.
- Ramachandram, D. and Taylor, G. W. (2017). Deep multimodal learning: A survey on recent advances and trends. *IEEE Signal Process Mag*, 34(6), pp. 96–108.
- Reaz, M. B. I., Hussain, M. S. and Mohd-Yasin, F. (2006). Techniques of EMG signal analysis: detection, processing, classification and applications. *Biol Proced Online*, 8(1), pp. 11–35.
- Reineking, T. (2014). *Belief Functions: Theory and Algorithms*. Ph.D. thesis, University of Bremen.
- Rezazadeh, I. M., Firoozabadi, M., Hu, H. and Golpayegani, S. M. R. H. (2012). Co-adaptive and affective human-machine interface for improving training performances of virtual myoelectric forearm prosthesis. *IEEE Trans Affect Comput*, 3(3), pp. 285–297.
- Robertson, J. W., Englehart, K. B. and Scheme, E. J. (2019). Effects of confidence-based rejection on usability and error in pattern recognition-based myoelectric control. *IEEE J Biomed Health Inform*, 23(5), pp. 2002–2008.
- Rodríguez-Carreño, Gila-Useros, L. and Malanda-Trigueros, A. (2012). Motor unit action potential duration: Measurement and significance. In I. Ajeena, ed., *Advances in Clinical Neurophysiology*, IntechOpen, chap. 7, p. 1.

- Roeleveld, K., Stegeman, D. F., Vingerhoets, H. M. and Oosterom, A. V. (1997). Motor unit potential contribution to surface electromyography. *Acta Physiol Scand*, 160(2), pp. 175–83.
- Rojas-Martínez, M., Mañanas, M. A. and Alonso, J. F. (2012). High-density surface EMG maps from upper-arm and forearm muscles. *J Neuroeng and Rehabil*, 9(1), 85.
- Saito, T. and Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS One*, 10(3), e0118432.
- Samadani, A. A. and Kulić, D. (2014). Hand gesture recognition based on surface electromyography. *36th Conf Proc IEEE Eng Med Biol Soc*, pp. 4196–4199.
- Saponas, T. S., Tan, D. S., Morris, D. and Balakrishnan, R. (2008). Demonstrating the feasibility of using forearm electromyography for muscle-computer interfaces. In *CHI '08 Proceedings of Physiological Sensing for Input*, vol. 1 and 2, p. 515–524.
- Savur, C. and Sahin, F. (2016). American sign language recognition system by using surface EMG signal. In *SMC, IEEE*, pp. 2872–2877.
- Scheme, E. J. and Englehart (2013). Validation of a selective ensemble-based classification scheme for myoelectric control using a three-dimensional fitts' law test. *IEEE Trans Neural Syst Rehabil Eng*, 21(4), pp. 616–623.
- Scheme, E. J. and Englehart, K. B. (2011). Electromyogram pattern recognition for control of powered upper-limb prostheses: state of the art and challenges for clinical use. *J Rehabil Res Dev*, 48(6), pp. 643–659.
- Scheme, E. J. and Englehart, K. B. (2015). A comparison of classification based

- confidence metrics for use in the design of myoelectric control systems. In *EMBC*, IEEE, pp. 7278–7283.
- Scheme, E. J., Hudgins, B. S. and Englehart, K. B. (2013). Confidence-based rejection for improved pattern recognition myoelectric control. *IEEE Trans Biomed Eng*, 60(6), pp. 1563–1570.
- Scheme, E. J., Fougner, A. L., Stavadahl, O., Chan, A. D. C. and Englehart, K. B. (2010). Examining the adverse effects of limb position on pattern recognition based myoelectric control. *Conf Proc IEEE Eng Med Biol Soc*, pp. 6337–6340.
- Scherberger, H. (2009). Neural control of motor prostheses. *Curr Opin Neurobiol*, 19(6), pp. 629–633.
- Schuetze, S. M. (1983). The discovery of the action potential. *Trends Neurosci*, 6, pp. 164–168.
- Sears, H. H. and Shaperman, J. (1991). Proportional myoelectric hand control: An evaluation. *Am J Phys Med Rehabil*, 70, pp. 20–28.
- Sensoy, M., Kaplan, L. M. and Kandemir, M. (2018). Evidential deep learning to quantify classification uncertainty. In *NeurIPS*, pp. 3183–3193.
- Shen, S., Gu, K., Chen, X., Yang, M. and Wang, R. (2019). Movements classification of multi-channel sEMG based on CNN and stacking ensemble learning. *IEEE Access*, 7, pp. 137489–137500.
- Sokolova, M. and Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Inf Process Manag*, 45(4), pp. 427–437.
- Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *J R Stat Soc Series B Stat Methodol*, 36(2), pp. 111–133.

- Su, J., Vargas, D. V. and Sakurai, K. (2019). One pixel attack for fooling deep neural networks. *IEEE Trans Evol Comput*, 23(5), pp. 828–841.
- Sun, W. et al. (2019). sEMG-based hand-gesture classification using a generative flow model. *Sensors*, 19(8), 1952.
- Szegedy, C. et al. (2014). Intriguing properties of neural networks. In *ICLR (Poster)*, pp. 1–10, <http://arxiv.org/abs/1312.6199>.
- Tatarian, K., Couceiro, M., Ribeiro, E. and Faria, D. (2018). Stepping-stones to transhumanism: An EMG-controlled low-cost prosthetic hand for academia. In *2018 International Conference on Intelligent Systems (IS)*, pp. 807–812.
- Tavakoli, M., Benussi, C., Lopes, P., Osorio, L. B. and de Almeida, A. T. (2018). Robust hand gesture recognition with a double channel surface EMG wearable armband and SVM classifier. *Biomed Signal Process Control*, 46, pp. 121–130.
- Thongpanja, S., Phinyomark, A., Phukpattaranont, P. and Limsakul, C. (2013). Mean and median frequency of EMG signal to determine muscle force based on time-dependent power spectrum. *J Electr Electron*, 19(3), pp. 51–56.
- Tommasi, T., Orabona, F., Castellini, C. and Caputo, B. (2013). Improving control of dexterous hand prostheses using adaptive learning. *IEEE Trans Robot*, 29(1), pp. 207–219.
- Trigili, E. et al. (2019). Detection of movement onset using EMG signals for upper-limb exoskeletons in reaching tasks. *J Neuroeng Rehabil*, 16(1), 45.
- Tsinganos, P., Cornelis, B., Cornelis, J., Jansen, B. and Skodras, A. (2019). Improved gesture recognition based on sEMG signals and TCN. In *ICASSP 2019*, pp. 1169–1173.
- Varma, S. and Simon, R. (2006). Bias in error estimation when using cross-validation for model selection. *BMC Bioinform*, 7, 91.

- Vidovic, M. M.-C. et al. (2016). Improving the robustness of myoelectric pattern recognition for upper limb prostheses by covariate shift adaptation. *IEEE Trans Neural Sys & Rehab Eng*, 24(9), pp. 961–970.
- Wahid, M. F., Tafreshi, R. and Langari, R. (2020). A multi-window majority voting strategy to improve hand gesture recognition accuracies using electromyography signal. *IEEE Trans Neural Syst Rehabil Eng*, 28(2), pp. 427–436.
- Wahid, M. F., Tafreshi, R., Al-Sowaidi, M. and Langari, R. (2018). Subject-independent hand gesture recognition using normalization and machine learning algorithms. *J Comput Sci*, 27, pp. 69–76.
- Wan, B. et al. (2010). Study on fatigue feature from forearm sEMG signal based on wavelet analysis. In *2010 IEEE ROBIO*, IEEE, pp. 1229–1232.
- Wang, J., Tang, L. and Bronlund, J. E. (2013). Surface EMG signal amplification and filtering. *Int J Comput Appl*, 82, pp. 15–22.
- Wei, W. et al. (2019). Surface-electromyography-based gesture recognition by multi-view deep learning. *IEEE Trans Biomed Eng*, 66(10), pp. 2964–2973.
- Widmaier, E. P., Raff, H. and Strang, K. T. (2010). *Vander’s Human Physiology: The Mechanisms of Body Function*, New York: McGraw Hill, chap. Muscle. 12th edn., pp. 250–291.
- Winkel, J. and Jørgensen, K. (1991). Significance of skin temperature changes in surface electromyography. *Eur J Appl Physiol Occup Physiol*, 63(5), pp. 345–348.
- Wu, L., Zhang, X., Zhang, X., Chen, X. and Chen, X. (2021). Metric learning for novel motion rejection in high-density myoelectric pattern recognition. *Knowl Based Syst*, 107165.

- Wu, S. and Nagahashi, H. (2014). Parameterized AdaBoost: Introducing a parameter to speed up the training of real AdaBoost. *IEEE Signal Process Lett*, 21(6), pp. 687–691.
- Yamamoto, Y. and Yamamoto, T. (1978). Dispersion and correlation of the parameters for skin impedance. *Med Biol Eng Comput*, 16(5), pp. 592–594.
- Yang, D., Zhao, J., Jiang, L. and Liu, H. (2012). Dynamic hand motion recognition based on transient and steady-state EMG signals. *Int J Humanoid Robotics*, 9(1), 1250007.
- Yang, D., Gu, Y., Jiang, L., Osborn, L. and Liu, H. (2017a). Dynamic training protocol improves the robustness of PR-based myoelectric control. *Biomed Signal Process Control*, 31, pp. 249–256.
- Yang, X., Chen, X., Cao, X., Wei, S. and Zhang, X. (2017b). Chinese sign language recognition based on an optimized tree-structure framework. *IEEE J Biomed Health Inform*, 21(4), pp. 994–1004.
- Young, A. J., Hargrove, L. J. and Kuiken, T. A. (2011). The effects of electrode size and orientation on the sensitivity of myoelectric pattern recognition systems to electrode shift. *IEEE Trans Biomed Eng*, 58(9), pp. 2537–2544.
- Young, A. J., Hargrove, L. J. and Kuiken, T. A. (2012). Improving myoelectric pattern recognition robustness to electrode shift by changing interelectrode distance and electrode configuration. *IEEE Trans Biomed Eng*, 59(3), pp. 645–652.
- Zanghieri, M. et al. (2020a). Robust real-time embedded EMG recognition framework using temporal convolutional networks on a multicore IoT processor. *IEEE Trans Biomed Circuits Syst*, 14(2), pp. 244–256.

- Zanghieri, M., Benatti, S., Conti, F., Burrello, A. and Benini, L. (2020b). Temporal variability analysis in sEMG hand grasp recognition using temporal convolutional networks. In *2020 2nd IEEE International Conference on AICAS*, pp. 228–232.
- Zhao, X., Chen, F., Hu, S. and Cho, J. (2020). Uncertainty aware semi-supervised learning on graph data. In *NeurIPS*, vol. 33, pp. 12827–12836.
- Zhu, X., Liu, J., Zhang, D., Sheng, X. and Jiang, N. (2017). Cascaded adaptation framework for fast calibration of myoelectric control. *IEEE Trans Neural Sys & Rehab Eng*, 25(3), pp. 254–264.
- Zia ur Rehman, M. et al. (2018a). Multiday EMG-based classification of hand motions with deep learning techniques. *Sensors*, 18(8), 2497.
- Zia ur Rehman, M. et al. (2018b). Stacked sparse autoencoders for EMG-based classification of hand motions: A comparative multi day analyses between surface and intramuscular EMG. *Appl Sci*, 8(7), 1126.

Appendix A

Hand movements demonstrations




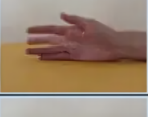
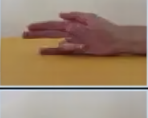
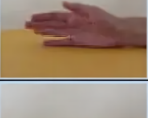






Exercise A		
1	Index flexion	
2	Index extension	
3	Middle flexion	
4	Middle extension	
5	Ring flexion	
6	Ring extension	
7	Little finger flexion	
8	Little finger extension	
9	Thumb adduction	
10	Thumb abduction	
11	Thumb flexion	
12	Thumb extension	

Figure 25: The 12 basic finger movements considered in the NinaPro Database 5, Exercise A, with corrected labels (Atzori and Müller 2015).


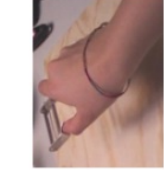






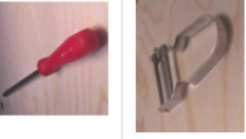



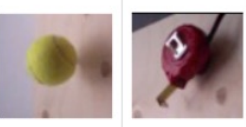
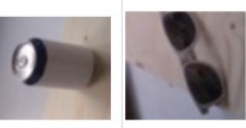
Description	Large Diameter	Adducted Thumb	Index Finger Extension	Medium Wrap	Writing Tripod	Power Sphere	Precision Sphere
Grasp							
Objects							

Figure 26: The 7 hand grasps with their relative holding objects considered in the NinaPro Database 6 (Palermo et al. 2017).

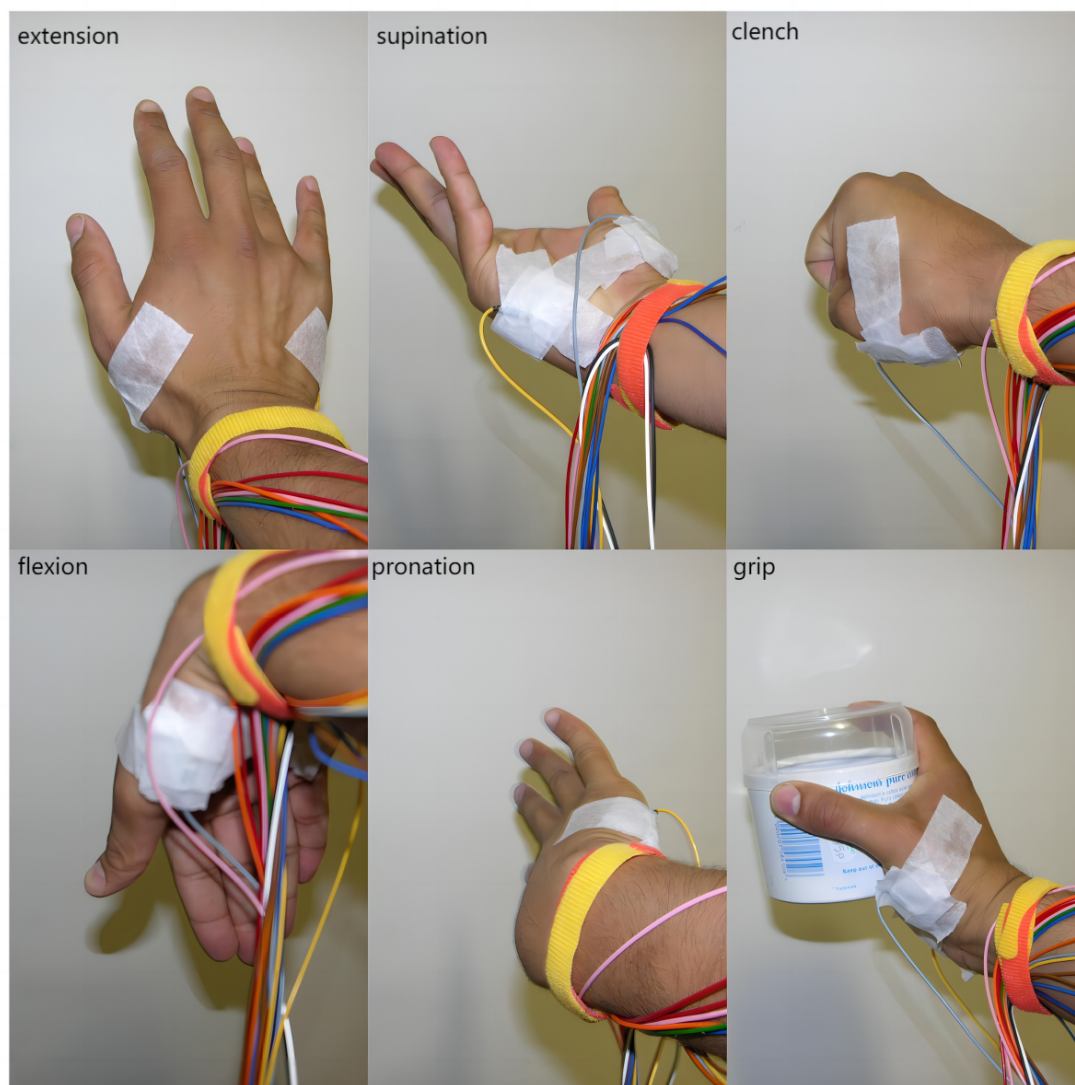


Figure 27: A subset of hand gesture examples from G. Dataset (Li et al. 2011).

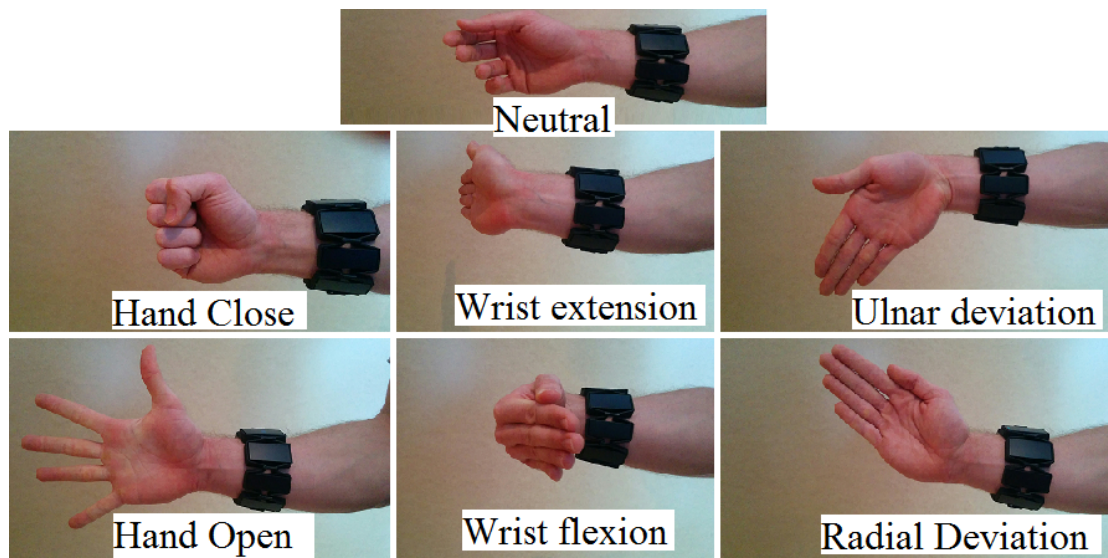


Figure 28: The 7 commonly seen hand gestures considered in the M. Dataset (Côté-Allard et al. 2019).

Appendix B

**The Determined Optimal
Hyperparameters of Models
employed for reliability analysis
of finger movement recognition**

Table 16: The Determined Optimal Hyperparameters of CNN with Stratified Nested Cross-Validation on *NinaPro DB5*

Subj	Outer Loop Fold	batch size	learning rate	optimizer
1	1	256	0.00861312	RMSprop
	2	128	0.00215868	RMSprop
	3	256	0.00143892	RMSprop
	4	128	0.00224912	ADAM
	5	128	0.00790112	ADAM
	6	128	0.00147652	RMSprop
2	1	128	0.00872215	RMSprop
	2	128	0.00288572	ADAM
	3	128	0.00569556	RMSprop
	4	128	0.00196093	ADAM
	5	128	0.00348368	ADAM
	6	128	0.00130826	RMSprop
3	1	128	0.00669941	RMSprop
	2	128	0.00385254	ADAM
	3	128	0.00375963	ADAM
	4	128	0.00159470	ADAM
	5	128	0.00552542	RMSprop
	6	128	0.00180949	ADAM
4	1	128	0.00417521	RMSprop
	2	256	0.00628941	RMSprop
	3	128	0.00446880	RMSprop
	4	128	0.00566787	RMSprop
	5	128	0.00551087	RMSprop
	6	128	0.00717378	RMSprop
5	1	256	0.00374168	RMSprop
	2	128	0.00370201	ADAM
	3	128	0.00504043	ADAM
	4	128	0.00169821	RMSprop
	5	128	0.00338842	ADAM
	6	128	0.00554601	RMSprop
6	1	128	0.00618573	ADAM
	2	256	0.00863308	RMSprop
	3	128	0.00777292	ADAM
	4	128	0.00848694	RMSprop
	5	128	0.00248306	ADAM
	6	128	0.00427281	ADAM
7	1	128	0.00176621	RMSprop
	2	128	0.00407085	ADAM
	3	128	0.00293482	RMSprop
	4	256	0.00200357	RMSprop
	5	128	0.00206999	RMSprop
	6	128	0.00262685	RMSprop
8	1	128	0.00147638	RMSprop
	2	128	0.00238904	RMSprop
	3	128	0.00180673	ADAM
	4	128	0.00208326	RMSprop
	5	128	0.00162371	RMSprop
	6	128	0.00214145	RMSprop
9	1	128	0.00558604	RMSprop
	2	128	0.00150895	RMSprop
	3	256	0.00707605	RMSprop
	4	256	0.00610104	RMSprop
	5	256	0.00246841	RMSprop
	6	128	0.00237642	RMSprop
10	1	128	0.00327037	RMSprop
	2	128	0.00181628	RMSprop
	3	128	0.00295268	ADAM
	4	128	0.00246366	ADAM
	5	128	0.00299895	RMSprop
	6	128	0.00102180	RMSprop

Table 17: The Determined Optimal Hyperparameters of ECNN-A with Stratified Nested Cross-Validation on *NinaPro DB5*

Subj	Fold	batch size	learning rate	optimizer	evidence fun
1	1	128	0.00695823	ADAM	ReLU
	2	128	0.00997057	ADAM	SoftPlus
	3	128	0.00874911	RMSprop	SoftPlus
	4	128	0.00335200	RMSprop	ReLU
	5	256	0.00913189	RMSprop	SoftPlus
	6	128	0.00217550	ADAM	ReLU
2	1	128	0.00878374	RMSprop	SoftPlus
	2	256	0.00696920	RMSprop	SoftPlus
	3	128	0.00943700	ADAM	SoftPlus
	4	256	0.00255285	ADAM	SoftPlus
	5	256	0.00492863	RMSprop	SoftPlus
	6	128	0.00419251	ADAM	ReLU
3	1	256	0.00740250	RMSprop	SoftPlus
	2	256	0.00952314	RMSprop	SoftPlus
	3	256	0.00642838	RMSprop	ReLU
	4	128	0.00701040	RMSprop	SoftPlus
	5	256	0.00400770	RMSprop	SoftPlus
	6	128	0.00662747	RMSprop	SoftPlus
4	1	128	0.00791172	RMSprop	SoftPlus
	2	128	0.00356949	RMSprop	SoftPlus
	3	128	0.00761020	RMSprop	SoftPlus
	4	128	0.00369350	ADAM	SoftPlus
	5	256	0.00495729	RMSprop	ReLU
	6	256	0.00442420	RMSprop	SoftPlus
5	1	128	0.00714004	ADAM	SoftPlus
	2	128	0.00987980	ADAM	SoftPlus
	3	128	0.00510656	ADAM	ReLU
	4	128	0.00176210	RMSprop	SoftPlus
	5	128	0.00203610	RMSprop	SoftPlus
	6	128	0.00677087	ADAM	SoftPlus
6	1	128	0.00459300	RMSprop	SoftPlus
	2	128	0.00336792	ADAM	SoftPlus
	3	128	0.00998516	ADAM	SoftPlus
	4	128	0.00217437	ADAM	SoftPlus
	5	128	0.00279520	RMSprop	ReLU
	6	128	0.00438666	ADAM	SoftPlus
7	1	128	0.00327256	SoftPlus	SoftPlus
	2	128	0.00988837	ADAM	SoftPlus
	3	128	0.00620641	ADAM	ReLU
	4	128	0.00463151	RMSprop	SoftPlus
	5	128	0.00577810	ADAM	SoftPlus
	6	128	0.00986300	ADAM	SoftPlus
8	1	128	0.00392580	ADAM	SoftPlus
	2	128	0.00304193	ADAM	ReLU
	3	256	0.00683400	ADAM	SoftPlus
	4	128	0.00463151	ADAM	SoftPlus
	5	128	0.00102860	RMSprop	SoftPlus
	6	128	0.00631100	RMSprop	SoftPlus
9	1	128	0.00421213	RMSprop	SoftPlus
	2	128	0.00635812	RMSprop	SoftPlus
	3	128	0.00249564	ADAM	SoftPlus
	4	256	0.00267723	RMSprop	SoftPlus
	5	128	0.00433935	ADAM	ReLU
	6	128	0.00380242	ADAM	SoftPlus
10	1	128	0.00384500	RMSprop	SoftPlus
	2	128	0.00318350	RMSprop	SoftPlus
	3	128	0.00287494	ADAM	ReLU
	4	256	0.00340249	ADAM	SoftPlus
	5	256	0.00225589	RMSprop	ReLU
	6	128	0.00367486	ADAM	SoftPlus

Table 18: The Determined Optimal Hyperparameters of ECNN-B with Stratified Nested Cross-Validation on *NinaPro DB5*

Subject	Outer Loop Fold	batch size	learning rate	optimizer	evidence fun	annealing step
1	1	128	0.00755720	ADAM	ReLU	60
	2	256	0.00937702	RMSprop	SoftPlus	55
	3	128	0.00993340	RMSprop	SoftPlus	60
	4	128	0.00610603	ADAM	SoftPlus	60
	5	128	0.00588229	RMSprop	SoftPlus	55
	6	128	0.00589500	RMSprop	SoftPlus	60
2	1	128	0.00718230	ADAM	SoftPlus	55
	2	128	0.00630821	RMSprop	SoftPlus	60
	3	128	0.00759272	ADAM	SoftPlus	60
	4	256	0.00182804	RMSprop	ReLU	60
	5	128	0.00569176	ADAM	SoftPlus	60
	6	128	0.00853500	ADAM	SoftPlus	60
3	1	128	0.00970321	ADAM	SoftPlus	60
	2	128	0.00764483	ADAM	ReLU	60
	3	128	0.00866606	RMSprop	SoftPlus	60
	4	256	0.00763013	RMSprop	SoftPlus	60
	5	128	0.00758694	ADAM	ReLU	55
	6	128	0.00841684	RMSprop	SoftPlus	60
4	1	128	0.00987511	ADAM	SoftPlus	60
	2	128	0.00899087	ADAM	SoftPlus	60
	3	128	0.00992393	ADAM	SoftPlus	60
	4	128	0.00851400	ADAM	SoftPlus	60
	5	128	0.00575878	RMSprop	SoftPlus	60
	6	128	0.00889460	RMSprop	SoftPlus	60
5	1	128	0.00804827	RMSprop	SoftPlus	60
	2	128	0.00701815	RMSprop	SoftPlus	60
	3	128	0.00844130	ADAM	SoftPlus	45
	4	256	0.00844045	RMSprop	SoftPlus	60
	5	128	0.00842967	ADAM	SoftPlus	45
	6	128	0.00877795	RMSprop	SoftPlus	60
6	1	128	0.00834386	RMSprop	SoftPlus	55
	2	128	0.00839500	ADAM	SoftPlus	60
	3	128	0.00727211	RMSprop	SoftPlus	60
	4	256	0.00669182	RMSprop	SoftPlus	60
	5	128	0.00762575	ADAM	SoftPlus	60
	6	128	0.00663232	RMSprop	SoftPlus	60
7	1	128	0.00816471	ADAM	SoftPlus	60
	2	128	0.00897314	ADAM	SoftPlus	55
	3	256	0.00998654	RMSprop	SoftPlus	60
	4	128	0.00905298	RMSprop	SoftPlus	60
	5	128	0.00992728	ADAM	SoftPlus	55
	6	128	0.00804264	RMSprop	SoftPlus	60
8	1	128	0.00412006	RMSprop	SoftPlus	55
	2	128	0.00534277	RMSprop	SoftPlus	60
	3	128	0.00346718	RMSprop	SoftPlus	60
	4	128	0.00543464	RMSprop	SoftPlus	60
	5	128	0.00989251	ADAM	SoftPlus	60
	6	256	0.00796879	RMSprop	SoftPlus	55
9	1	256	0.00731462	RMSprop	SoftPlus	60
	2	128	0.00739066	ADAM	SoftPlus	60
	3	128	0.00710547	ADAM	SoftPlus	55
	4	128	0.00984175	ADAM	SoftPlus	55
	5	128	0.00819889	RMSprop	SoftPlus	60
	6	256	0.00799579	ADAM	SoftPlus	60
10	1	128	0.00896102	ADAM	SoftPlus	55
	2	128	0.00805946	RMSprop	SoftPlus	60
	3	128	0.00898895	RMSprop	SoftPlus	60
	4	128	0.00840827	RMSprop	SoftPlus	60
	5	256	0.00701282	RMSprop	SoftPlus	60
	6	128	0.00573632	RMSprop	SoftPlus	60

Table 19: The Determined Optimal Hyperparameters of ECNN-C with Stratified Nested Cross-Validation on *NinaPro DB5*

Subject	Outer Loop Fold	batch size	learning rate	optimizer	evidence fun	tau
1	1	128	0.00292867	RMSprop	SoftPlus	0.0104
	2	128	0.00711669	ADAM	SoftPlus	0.0100
	3	128	0.00827213	ADAM	ReLU	0.0100
	4	256	0.00111971	ADAM	ReLU	0.0102
	5	256	0.00194469	RMSprop	ReLU	0.0104
	6	128	0.00219270	RMSprop	ReLU	0.0158
2	1	128	0.00469811	RMSprop	SoftPlus	0.0108
	2	128	0.00648811	RMSprop	SoftPlus	0.0155
	3	128	0.00335853	RMSprop	SoftPlus	0.0105
	4	256	0.00762505	ADAM	SoftPlus	0.0115
	5	256	0.00613626	ADAM	ReLU	0.0104
	6	256	0.00103725	ADAM	SoftPlus	0.0185
3	1	128	0.00999356	ADAM	SoftPlus	0.0100
	2	128	0.00255055	ADAM	SoftPlus	0.0101
	3	128	0.00799018	ADAM	SoftPlus	0.0104
	4	256	0.00351461	ADAM	ReLU	0.0102
	5	256	0.00746422	RMSprop	SoftPlus	0.0100
	6	128	0.00969107	RMSprop	SoftPlus	0.0101
4	1	256	0.00255049	RMSprop	ReLU	0.0106
	2	256	0.00543337	ADAM	ReLU	0.0101
	3	256	0.00400591	ADAM	ReLU	0.0101
	4	256	0.00237541	ADAM	SoftPlus	0.0102
	5	128	0.00543133	RMSprop	SoftPlus	0.0104
	6	128	0.00244918	ADAM	SoftPlus	0.0119
5	1	256	0.00555642	RMSprop	SoftPlus	0.0115
	2	256	0.00209180	ADAM	Exp	0.0117
	3	256	0.00285707	RMSprop	ReLU	0.0106
	4	128	0.00134497	ADAM	ReLU	0.0195
	5	128	0.00229343	ADAM	ReLU	0.0101
	6	256	0.00149723	RMSprop	ReLU	0.0143
6	1	128	0.00155251	RMSprop	ReLU	0.0100
	2	256	0.00161771	RMSprop	ReLU	0.0103
	3	128	0.00407107	ADAM	SoftPlus	0.0110
	4	128	0.00402627	ADAM	SoftPlus	0.0102
	5	256	0.00590821	ADAM	Exp	0.0111
	6	256	0.00435864	ADAM	ReLU	0.0112
7	1	256	0.00216571	RMSprop	SoftPlus	0.0103
	2	128	0.00205496	RMSprop	ReLU	0.0103
	3	256	0.00148712	ADAM	ReLU	0.0109
	4	256	0.00431477	ADAM	SoftPlus	0.0102
	5	256	0.00263851	ADAM	ReLU	0.0103
	6	256	0.00135566	ADAM	ReLU	0.0167
8	1	128	0.00831896	RMSprop	SoftPlus	0.0101
	2	256	0.00502280	RMSprop	SoftPlus	0.0106
	3	256	0.00328582	ADAM	SoftPlus	0.0103
	4	128	0.00254655	ADAM	ReLU	0.0100
	5	256	0.00546814	RMSprop	SoftPlus	0.0116
	6	256	0.00482164	ADAM	Exp	0.0102
9	1	256	0.00655349	ADAM	SoftPlus	0.0102
	2	256	0.00297165	ADAM	Exp	0.0121
	3	128	0.00163260	RMSprop	ReLU	0.0185
	4	256	0.00101235	RMSprop	ReLU	0.0145
	5	128	0.00456503	ADAM	ReLU	0.0102
	6	256	0.00346760	ADAM	ReLU	0.0103
10	1	128	0.00368632	ADAM	SoftPlus	0.0101
	2	256	0.00940509	ADAM	SoftPlus	0.0102
	3	256	0.00276342	ADAM	ReLU	0.0111
	4	256	0.00163194	ADAM	SoftPlus	0.0102
	5	256	0.00525047	ADAM	Exp	0.0102
	6	256	0.00885823	RMSprop	SoftPlus	0.0106