

Electromyography-based Biometrics for Secure and Robust Personal Identification and Authentication

by

Ashirbad Pradhan

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Systems Design Engineering

Waterloo, Ontario, Canada, 2024

© Ashirbad Pradhan 2024

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Member: Dr. Wei Chen
Professor, Dept. Biomedical Engineering,
University of Sydney

Supervisor(s): Dr. Ning Jiang
Professor, Dept. of Systems Design Engineering, University of Waterloo

Dr. James Tung
Professor, Dept. of Systems Design Engineering, University of Waterloo

Internal Members: Dr. Shi Cao
Associate Professor, Dept. of Systems Design Engineering,
University of Waterloo

Dr. Richard Nuckols
Assistant Professor, Dept. of Systems Design Engineering,
University of Waterloo

Internal-External Member: Dr. George Shaker
Associate Professor, Dept. of Electrical and Computer Engineering,
University of Waterloo

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

Important contributions towards this thesis were made by my supervisors Dr. Ning Jiang and Dr. James Tung. Their input and guidance aided towards the overall completion of the thesis work.

Chapter 3 and 4 of this thesis are published papers co-authored by myself, Dr. Jiayuan He and Dr. Jiang. For these chapters, I developed the methodology and experimental design, analyzed the data, and wrote the papers with guidance and support from Dr. Jiang and Dr. He. The data collection for these studies was performed by Dr. He.

Chapter 5 of this thesis is a published paper co-authored by myself, Dr. He and Dr. Jiang. For this chapter, I developed the methodology and experimental design, collected the data, analyzed it and wrote the papers with guidance and support from Dr. Jiang and Dr. He. Additionally, I contributed in making the dataset and codebase open access with the support of Dr. He and Dr. Jiang.

Chapter 6 of this thesis is a published paper co-authored by myself, Dr. He, Hyowon Francis Lee, and Dr. Jiang. For this chapter, I developed the methodology and experimental design, analyzed the data, and wrote the papers with guidance and support from Dr. Jiang and Dr. He. Hyowon contributed in the identification analysis (Figure 6.6 and Figure 6.7) and in the overall writing of the paper.

Chapter 7 of this thesis contains unpublished materials which are currently submitted to a journal and under review. The paper is co-authored by myself, Dr. He, Dr. Tung, and Dr. Jiang. For this chapter, I developed the methodology and experimental design, analyzed the data, and wrote the papers with guidance and support from Dr. Jiang, Dr. Tung, and Dr. He. Dr. Tung, and Dr. Jiang contributed in the reviewing of the manuscript.

I certify, with the above qualifications, that this thesis and the research to which it refers is the product of my own work.

Abstract

Recently, electromyogram (EMG), the electrical activity of skeletal muscles, has been proposed as a novel biometric trait to address the limitations of current biometrics, such as fingerprint and facial recognition. A unique property of EMG as a biometric trait is that it allows for distinguishable patterns from different limb movements (e.g. hand gestures), enabling individuals to set personalized passwords comprising multiple gestures for dual-security systems, i.e., both biometric-level and password-level. This is fundamentally different from other physiological signals such as electrocardiogram (ECG) and electroencephalogram (EEG), which are highly difficult for the user to voluntarily control with sufficient precision. This unique advantage has facilitated EMG-based biometrics for two different applications: authentication, where a user can access personal devices, and identification, where the system determines the closest match within a database. To establish EMG as a novel biometric trait, the following two properties need to be thoroughly investigated: 1) the ability to accurately detect the genuine user from all the other users (uniqueness), and 2) retaining the biometric performance over multiple sessions and multiple days (robustness). The overarching aim of this PhD research is to investigate these properties by addressing a series of research questions in the following studies.

In the first study (Chapter 3), the effect of EMG system parameters such as the feature extraction methods and the number of channels are investigated for improved biometric performance. Three robust feature extraction methods, Time-domain (TD), Frequency Division Technique (FDT), and Autoregressive (AR) features, and their combinations were investigated, while the number of channels varied from one to eight. The results showed that for all the feature extraction methods, the performance of a four-channel setup plateaued with a further increase in channels. For a four-channel system, the authentication performance resulted in an average equal error rate (*EER*) of 0.04 for TD features, 0.053 for FDT features, and 0.10 for AR features. The identification mode resulted in an average *Rank-1 accuracy* was 97% for TD features, 87.6% for FDT features, and 63.7% for AR features. Thus, combining the TD feature set and a four-channel EMG is recommended for optimal biometric performance.

In the second study (Chapter 4), the dual-security property of EMG is facilitated by the development of a multi-code framework. Such a framework allows the combination of hand gestures to form an access code. In this study, three levels of fusion, score, rank, and decision were investigated for the two biometric applications. The biometric performance of the fusion schemes was analyzed while varying the codelength from one to six. For a codelength of four, the authentication *EER* was 0.006 using a decision-level fusion scheme using a weighted majority voting. For the identification mode, the score-level fusion scheme resulted in a *Rank-2 accuracy* of 99.9% for a codelength of four. The multi-code biometric system provided improved dual-mode security based on the personalized codes and biometric traits of individuals.

However, the above two studies and the majority of the current EMG-based biometric research face two critical limitations: 1) a small subject pool, comparative to other more established biometric traits, and 2) single-session data sets. In multi-day scenarios, there is performance degradation of EMG-based biometrics. In the third study (Chapter 5) a multi-day and large-sample dataset collection was performed to address these limitations. For the research study, EMG data was collected from 43 participants over three different days with long separation (Days 1, 8, and 29) while performing 16 different static hand/wrist gestures with seven repetitions. The dataset was made public as the GRABMyo dataset. In study four (Chapter 6), a multi-day analysis involving training data and testing data from different days of the GRABMyo dataset was employed to test the robustness of the EMG-based biometrics in practical scenarios. The cross-day authentication using the FDT features extraction resulted in a median *EER* of 0.039 when the code (gestures) was secure, and an *EER* of 0.068 when the code (gestures) was leaked to intruders. The cross-day identification achieved a median *rank-5 accuracy* of 93.0%.

For improving multi-day performance, robust feature extraction methods that employ deep learning are warranted. In study five (Chapter 7), a convolutional feature engineering method, MyoBM-Net, is proposed. It involves a two-stage training paradigm for improving the authentication performance. In a cross-day analysis, the MyoBM-Net resulted in a median *EER* of 0.003 and 0.008 when the gesture (code) is safe and compromised, respectively, thus suggesting superior performance than the traditional feature extraction method. The findings suggest that the performance of EMG-based biometrics is comparable to conventional biometrics for both authentication and identification applications.

The results show the potential of using EMG signals for biometric identification in real-world scenarios. The multi-code framework facilitates the combination of gestures as passcodes. The large multi-day dataset will support further research on EMG-based biometrics and other gesture recognition applications. The MyoBM-Net architecture will enable the development of new applications using the GRABMyo dataset, leading to accurate and robust biometric performance. This could lead to EMG-based biometrics being used as an alternative to traditional biometric methods.

Acknowledgements

Funding for this research was provided by the Natural Sciences and Engineering Research Council of Canada (NSERC-Discovery #072169). I was awarded the prestigious NSERC-Canada Graduate Scholarship that supported me throughout the PhD journey.

I would like to extend my deepest gratitude to my supervisors, Dr. Ning Jiang and Dr. James Tung. Dr. Jiang's promptness and willingness to devote invaluable time to my progress have been pivotal throughout my research journey. His passion for EMG based research has been a constant source of inspiration. Dr. Tung's unwavering dedication and support, especially during challenging moments has made this milestone a reality. I am sincerely thankful for both their ongoing support and guidance, which have greatly contributed to my development as a researcher.

I would like to extend my gratitude to the committee members for their invaluable feedback, that helped me complete my thesis. I would also like to thank Dr. Jiayuan He for sharing his expertise on EMG signal processing and assisting me throughout the PhD research. I would also like to thank all the lab mates in my research group, past and present, for always being available to answer questions and brainstorm with. Thank you members of the Waterloo Ebionics lab and NRE lab, I value every moment spent in the lab with you all.

I would like to thank the Department of Systems Design Engineering, the staff, and the graduate chair(s) without whose help I could not finish the PhD milestones. I would like to thank Prof. Jennifer Howcroft for the numerous TA opportunities and for sharing her professional knowledge. I would also like to thank Dr. Usha Kuruganti and Dr Victoria Chester, my masters' supervisors for extending their support and checking on my progress during my PhD journey.

I want to express my deepest gratitude to my partner Ana. Words cannot express her unwavering help, research-support, constant moral-support, and being there for me during my achievements as well as through tough times. I will always be indebted to her for the behind-the-scene support in numerous journal manuscripts and presentations.

I want to express my heartfelt gratitude to the amazing friends I have made during this incredible journey. The cherished memories with Waterloo friends from Westcourt and Keatsway sustained me through the challenges of the COVID period, and have played an invaluable role in my personal and professional growth. I am also grateful for the memories with friends from Albert St during this period. I am very grateful for the experiences with my SYDE friends who have shared similar journeys. I am thankful to my friends from New Brunswick who have always checked up on me and still make plans with me. My friends from NITR in Waterloo have always helped me unconditionally. I would also like to thank my friends from CLV; you have

been the most amazing neighbors. I am indebted to my friends from undergraduate, LUNATICS, for being the constant support system through the PhD journey.

Finally, this work would not have been possible without the constant support of my family. I am endlessly grateful to my parents for their unconditional love and care. Without their help, I could not have made the tough decisions in life. I want to express my gratitude to my cousin, Gyana, my closest family member in Canada and has always been there to help me. As a final thanks I would acknowledge all the 43 participants who helped me make GRABMyo dataset a reality. Many others, knowingly or unknowingly, have contributed to the successful completion of this work, and I extend my heartfelt thanks to them all

Table of Contents

Examining Committee	ii
Author's Declaration	iii
Statement of Contributions	iv
Abstract	v
Acknowledgements	vii
List of Figures	xii
List of Tables	xx
1 Introduction	1
1.1 Thesis Organization	4
2 Literature Review	5
2.1 Biometrics	5
2.2 Surface Electromyogram (EMG)	7
2.3 EMG-Biometrics	10
2.4 Multi-code framework	12
2.5 Multi-day Robustness of EMG-Biometrics	15

2.6	Convolutional feature engineering using EMG	17
2.7	Research Objectives and Hypothesis	19
3	Performance Optimization of Surface Electromyography Based Biometric Sensing System for Both Authentication and Identification	21
3.1	Chapter Introduction	21
3.2	Methods	23
3.3	Results	29
3.4	Discussion	33
3.5	Chapter Conclusions	38
4	Score, Rank, and Decision-Level Fusion Strategies of Multicode Electromyogram-Based Authentication and Identification Biometrics	39
4.1	Chapter Introduction	39
4.2	Methods	40
4.3	Results	51
4.4	Discussion	54
4.5	Chapter Conclusions	58
5	Multi-day Dataset of Forearm and Wrist Electromyogram for Hand Gesture Recognition and Biometrics	59
5.1	Chapter Introduction	59
5.2	Background & Summary	60
5.3	Methods	62
5.4	Data Records	64
5.5	Technical Validation	66
5.6	Usage Notes	71
5.7	Code Availability	72

6	Multi-day Analysis of Wrist Electromyogram-based Biometrics for Authentication and Personal Identification	73
6.1	Chapter Introduction	73
6.2	Methods	74
6.3	Results	83
6.4	Discussion	89
6.5	Chapter Conclusion	94
7	MyoBM-Net: Two-stage Convolutional Feature Engineering for Long-term Biometric Authentication using Wrist Electromyography	96
7.1	Chapter Introduction	96
7.2	Methods	97
7.3	Results	108
7.4	Discussion	112
7.5	Chapter Conclusion	116
8	Conclusions	118
8.1	Summary of Results	118
8.2	Limitations and Future Research	120
8.3	Practical Significance	123
	References	124
	Appendices	140

List of Figures

1.1	Flow diagram of the research studies	3
2.1	The framework of an EMG-based authentication system showing the three components: Possession, Knowledge, and Biometrics	6
2.2	EMG signals collected from electrodes placed on the forearm. There are two types of signal acquisition: monopolar and single-differential EMG	8
2.3	A) Single-code EMG Biometric system. The processed input is matched with the template(s) in the database and a score is generated for every matching operation. B) Multicode EMG biometric system combines individual codes using a post-matching fusion at one of the three levels—score, rank, and decision. For both systems, the ranking module is present only for identification, whereas the authentication application bypasses the ranking module, and a decision is made based on the score threshold.	14
2.4	The structure of the convolutional neural network used in a gesture recognition study [45]. The study used three convolutional blocks and one fully connected layer block to match the input (409 frames x 6 channels) and output (17 classes) .	18
3.1	Positions of the sixteen monopolar electrodes on the forearm (dorsal view). In each ring, eight electrodes (grey rectangles on the forearm) are evenly placed, and one electrode is placed in the centerline of the elbow crease. The distance between the centers of the two rings was 2 cm, and the proximal ring was placed one third of the forearm length between the elbow crease and the upper ring. . . .	23

3.2	Sixteen gesture classes investigated in the study: (A) lateral prehension (LP), (B) thumb adduction (TA), (C) thumb and little finger opposition (TLFO), (D) thumb and index finger opposition (TIFO), (E) thumb and little finger extension (TLFE), (F) thumb and index finger extension (TIFE), (G) index and middle finger extension (IMFE), (H) little finger extension (LFE), (I) index finger extension (IFE), (J) thumb extension (TE), (K) wrist flexion (WF), (L) wrist extension (WE), (M) forearm supination (FS), (N) forearm pronation (FP), (O) hand open (HO), (P) hand close (HC).	26
3.3	(A and B). The median <i>AUC</i> and the median <i>EER</i> for the three feature extraction methods (FDT, TD and AR) and number of channels (1-8) for the Normal Test (left column), Leaked Test (middle column) and Self Test (right column) scenario of the authentication system. Each line indicates the respective median values, and shaded area indicates the corresponding middle 50 percentile (25% to 75%). (C) The range of <i>EER</i> values (EER_r) for the feature extraction methods (FDT, TD and AR) and number of channels (1-8) for the Normal Test, Leaked Test, and Self Test scenario of the authentication system. The range is calculated as a part of the sequential forward selection algorithm at the end of every iteration.	30
3.4	(A) The median <i>RIE</i> and the median <i>R5E</i> for the processing methods (FDT, TD, AR) and the number of channels (1-8) for the identification system. Each line indicates the respective median values, and shaded area indicates the corresponding middle 50 percentile (25% to 75%). (C) The range of the <i>RIE</i> (RIE_r) and <i>R5E</i> ($R5E_r$) for the feature extraction methods (FDT, TD and AR) for different number of channels (1-8) for the identification system. The range is calculated as a part of the sequential forward selection algorithm at the end of every iteration.	31
3.5	(A and B) The median <i>RIE</i> and the median <i>R5E</i> for the processing methods (FDT, TD, AR), their combinations (TD, TD+FDT, TD+AR) and number of channels (1-8) for the identification system. Each line indicates the respective median values, and shaded area indicates the corresponding middle 50 percentile (25% to 75%). (C and D) The range of the <i>RIE</i> (RIE_r) and <i>R5E</i> ($R5E_r$) for the feature extraction methods (FDT, TD and AR) and the combined feature sets (TD, TD+FDT, TD+AR) for different number of channels (1-8) for the identification system. The range is calculated as a part of the sequential forward selection algorithm at the end of every iteration	34
3.6	<i>CMC</i> curve for the identification system using 4 channels EMG and TD Feature extraction.	35

4.1	A) Unicode EMG Biometric system. The processed input is matched with the template(s) in the database and a score is generated for every matching operation. B) Multicode EMG biometric system combines individual codes using a post-matching fusion at one of the three levels—score, rank, and decision. For both the systems, the ranking module is present only for identification, whereas the authentication application bypasses the ranking module, and a decision is made based on the score threshold.	41
4.2	Positions of the eight surface electrodes on the forearm (dorsal view, two electrodes are not visible as they are in the posterior side). These eight electrodes formed four bipolar pairs in the proximal-distal orientation and the differential EMG was used for subsequent signal processing. The first pair is placed on the flexor carpi ulnaris (FCU) muscle and the remaining three are placed uniformly around the forearm.	43
4.3	A) The genuine and imposter score sets as the output of a matcher. The anchor value is calculated based on the overlap region and the mean values of the genuine and imposter score set [105]. B) The identification involves sample by sample comparison of the presented user to the template database. The outcome contains a genuine score (blue), the rest are imposter scores (brown). The best case is when the genuine score shows up in the tail of the score distribution. The best score in the tail is (in case of a match) an outlier with respect to the Weibull distribution model of the imposter scores [52].	47
4.4	The working of a multimodal EMG identification mode. A scenario with five users and three codes are presented, and different fusion schemes are performed. In a rank-4 system, four users are granted access. In the case of less than four users qualifying the selection criteria, the remaining users are chosen randomly (as seen in Decision-AND, one of user D or user E is selected at random)	49
4.5	Mean <i>EER</i> for the multicode authentication mode scenarios (normal test and leaked test). The codelength (M) was varied from 1 to 6 and different fusion schemes were employed: Decision-AND, Decision-WM, Score-MM, and Score-Norm. The dashed error bars represent the standard deviation.	50
4.6	A) Mean <i>FAR</i> for varying codelength (M) and rank (K) of the multimodal identification mode using different fusion schemes. B) Mean <i>FRR</i> for varying codelength (M) and rank (K) of multimodal identification mode using different fusion schemes. All bars are color-coded according to the legend on the right. Note that the purpose of presentation clarity, the bars with values higher than 0.2 have been clipped at 0.2.	51

5.1	Electrode Positions and Gesture list. Fig 1A (left) shows the electrode positions used in the study. There are two groups of electrodes: forearm and wrist. Each group comprises of two rings (proximal and distal). Monopolar EMG was recorded from a total of 28 channels (16 Forearm and 12 wrist). Fig 1B (right) shows the Sixteen gesture classes investigated in the study (numbered 1 – 16): lateral prehension (LP), thumb adduction (TA), thumb and little finger opposition (TLFO), thumb and index finger opposition (TIFO), thumb and little finger extension (TLFE), thumb and index finger extension (TIFE), index and middle finger extension (IMFE), little finger extension (LFE), index finger extension (IFE), thumb extension (TE), wrist flexion (WF), wrist extension (WE), forearm supination (FS), forearm pronation (FP), hand open (HO), and hand close (HC). The REST gesture was also collected in each trial repetition.	61
5.2	Representative amplitude and power spectrum plots. Left Column: The amplitude values (in mV) for the four electrode rings: proximal and distal forearm and wrist (from top to bottom). Right Column: The fast fourier transform (FFT) power spectrum (in dB/Hz) for the four electrode rings: proximal and distal forearm and wrist (from top to bottom). The forearm rings constitute 8 channels each and the wrist rings constitute 6 channels each (shown in different colors).	67
5.3	SNR and CCN comparison plots. Panel A, i.e. the top row shows <i>SNR</i> values (dB), with each spoke representing one motion. Left: the forearm proximal and distal electrode setup for session 1-3 (added as suffix FP1, FP2, . . . ,FD3); middle: wrist proximal and distal electrode setup for session 1-3 (added as suffix WP1,WP2, . . . ,WD3); and (right) the Ninapro signals collected from the forearm for a single day. Panel B, i.e. the bottom row shows the <i>CCN</i> values for the forearm (left), wrist (middle) and Ninapro data (right) in a similar manner.	68
5.4	ROC curves for HGR and Biometric validation. The <i>ROC</i> curves are plotted for the HGR evaluation (A) and Biometric Evaluation (B) for the forearm and wrist data averaged for the multiple users (N=43), sessions (N=3) as well as for the Ninapro Data (#users = 40, #sessions = 1). The dotted line represents the point of intersection on the <i>ROC</i> curve from which the <i>EER</i> value is obtained.	70
6.1	Positions of the six electrodes on the wrist (dorsal view, three electrodes are on the posterior wrist surface). The electrodes were equally spaced around the wrist at a distance of one cm from the styloid process. The monopolar EMG of each ring was acquired for subsequent processing and analysis.	75

6.2	The normal test and the leaked test scenarios for biometric authentication. In both the scenarios the target user's identity is known to the claimant. The access code is unknown to the claimant in the normal test scenario, while it is known to the claimant in a leaked test scenario.	79
6.3	(From left to right) Within-day, Single cross-day and Cumulative cross-day analysis. The corresponding training (enrollment) data for each analysis are represented in green; the testing (claimant) data are represented in dark green (for within-day analysis) and blue (for cross-day analysis). The x-axis represents the timeline of the study consisting of multiple days; the x-axis tick marks represent the different trials performed in one day. Each of the analysis is repeated for the three days.	82
6.4	Biometric authentication and identification performance. For the multi-day authentication performance (Fig 4A), the single-code and multi-code <i>EER</i> values for the normal test and leaked test scenario are shown. The x-axis represents the three multi-day analyses: within-day (WD), single cross-day (SCD), and the cumulative cross-day (CCD) and the y-axis represents the performance error. Each boxplot represents the interquartile range (IQR, 25th – 75th percentile) and the center horizontal line represents the median <i>EER</i> value. The whiskers (solid vertical lines) represent the datapoints within the $1.5 \cdot \text{IQR}$ threshold. The outliers (solid black circles) are defined as those individuals whose <i>EER</i> was greater than the $1.5 \cdot \text{IQR}$ threshold. For the multi-day identification, the <i>rank-1</i> and <i>rank-5 accuracy</i> for the single-code and multi-code configuration are shown. The x-axis represents the three multi-day analyses, and the y-axis represents the rank-K accuracy where $K = 1, 5$	83
6.5	DET and CMC curves for the biometric tests. The averaged detection error DET curves for the two authentication scenarios: normal test (Fig 5A) and leaked test (Fig B), in three multi-day analyses: within-day (WD), single cross-day (SCD) and cumulative cross-day (CCD), and the two configurations: single-code and multi-code. The x-axis represents the FAR values, and the y-axis represents the FRR values. The averaged CMC curves (Fig 5B) are shown for the identification mode in the three multi-day analyses and the two configurations. The x-axis represents the Rank (K) of the identification system, and the y-axis represents the identification accuracy.	85

6.6	Individual identification performance. The <i>rank-5 accuracy</i> for all the users for the WD (top), SCD (middle) and CCD (bottom) analysis is illustrated. The multi-code values are shown in violet color and the corresponding single-code values are shown in green color.	86
6.7	Biometric identification performance. The <i>rank-1</i> , <i>rank-3</i> and <i>rank-5 accuracy</i> for the single-code and multi-code configuration are shown. The x-axis represents the three multi-day analyses, and the y-axis represents the rank-K accuracy where $K = 1, 5$. Each boxplot represents the interquartile range (IQR, 25th – 75th percentile) and the center horizontal line represents the median accuracy value. The whiskers (solid vertical lines) represent the datapoints within the $1.5 \times \text{IQR}$ threshold. The outliers (solid black circles) are defined as those individuals whose accuracy was greater than the $1.5 \times \text{IQR}$ threshold.	87
6.8	Intruder analysis for the identification mode. For the multi-day authentication performance, the single-code and multi-code <i>rank-5 accuracy</i> and <i>IRR</i> values are shown. The x-axis represents the three multi-day analyses: within-day (WD), single cross-day (SCD), and the cumulative cross-day (CCD) and the y-axis represents the <i>rank-5 accuracy</i> (left), and the <i>IRR</i> (right). Each boxplot represents the interquartile range (IQR, 25th – 75th percentile) and the center horizontal line represents the median <i>rank-5 accuracy/IRR</i> value.	88
7.1	Positions of the six electrodes on the wrist (dorsal view, three electrodes on the posterior wrist surface are not shown). The monopolar EMG of each ring was acquired for subsequent processing and analysis. The diagram is obtained from the data descriptor [107].	98
7.2	MyoBM-Net Convolutional Feature Extractor. Two network architectures (top-bottom) are proposed. The conv2d-based (C2D) utilized channel-specific information, while conv1d-based (C1D) used spatial information. Each architecture includes four blocks of convolution-pooling, normalization layers. The input consists of segmented EMG with 6 channels and 409 timepoints (corresponding to 250ms). The time-axis converges to a value of one (linear), which enables the extraction of relevant features. The output of the C2D and C1D architectures is a linear feature vector of length 36 and 32, respectively.	100

7.3	The two-stage training and authentication process. A) Twostage training paradigm for the MyoBM-Net feature extraction method: Stage 1 (S1) training to obtain a generalized feature extractor, and Stage 2 (S2) training for a user-specific authentication network. B) Authentication process involving feature extraction using the S2-trained MyoBM-Net, followed by biometric matching. The test features are compared to the enrollment features to generate a matching score, which is compared to a threshold value, and a final decision is generated.	103
7.4	Authentication performance comparison of network architectures and training paradigms. For the normal test (left) and leaked test (right), the <i>EER</i> values for SCD and CCD training are plotted. The x-axis represents the two network architectures: C2D and C1D, and the y-axis represents the performance error. The colors represent the three training paradigms: S1, S2-BL and S2. Each boxplot represents the interquartile range (IQR, 25th – 75th percentile) and the center horizontal line represents the median <i>EER</i> value. The whiskers (solid vertical lines) represent the datapoints within the $1.5 \cdot \text{IQR}$ threshold.	107
7.5	Authentication performance of S2-trained networks is compared with the traditional methods. For the normal test (left) and leaked test (right), the <i>EER</i> values for SCD and CCD training (x-axis) are plotted. The y-axis represents the performance error. The colors represent the feature extractors: FDT, C2D, and C1D. Box and whisker The boxes and whiskers are similar to Figure 7.4	109
7.6	Authentication performance of S2-trained C1D network in a CCD scenario for all the users. The <i>EER</i> values (y-axis) for the normal test (top) and leaked test (bottom) are plotted. The y-axis represents the 43 users in the database. The boxes and whiskers are presented similar to Figure 7.4	110
7.7	The feature space analysis of S1-trained networks. The t-SNE plots (1-3, from left) and DBI (right) values are plotted. The x-axis and yaxis for the t-SNE plots represent the reduced dimensions. The plots show the clustering of users' feature space for an identical code sequence in the three feature extraction methods (FDT, C2D and C1D). In the DBI plot (right), the x-axis denotes the training scenarios, and the y-axis denotes the DBI values. The colors represent the feature extractors: FDT, C2D, and C1D. The boxes are presented similar to Figure 7.4 . .	112
7.8	The feature space of S2 trained C1D network in a CCD testing for a sample true user. The t-SNE plot shows the visualization of feature space of a true user on two different days and their relative separation from the imposters.	113

D.1	The electrode locations on the wrist (12) and forearm (16) as used in the GRAB-Myo Dataset	145
-----	--	-----

List of Tables

2.1	Comparison of current biometric traits [25]	7
2.2	List of recent studies on EMG based biometric authentication and identification	11
2.3	Multi-day forearm and Wrist EMG Datasets	16
3.1	TD Feature Extraction and Four-Channel EMG Configuration	37
4.1	Recommended Parameter Values of Multimodal EMG Biometric System	56
5.1	Database Summary	63
5.2	Multi-Day EMG Database. MATLAB Codes: File Reading and Signal Processing	65
5.3	Mean ($\pm$ STD) HGR and Biometric performance.	69
6.1	Biometric Traits Characteristics	91
7.1	Network Hyperparameters	102
8.1	Factors affecting EMG-Biometrics performance	121
A.1	The unicode classification performance of each gesture for the Normal Test, Leaked Test and Identification scenarios.	141
B.1	Within Day Identification Performance of Gestures	142
C.1	GRABMyo Dataset Participant Demographics and Measurements	143

Chapter 1

Introduction

Biometrics has become an integral part of current security systems and has found application in consumer electronics, public security, and private security. These biological and behavioral traits have been utilized to identify an individual or verify an individual's identity. Conventional biometrics, such as fingerprints and facial scans, have been widely used in smartphones and laptops in our daily lives [143]. However, with technological advancements, there are increasing risks of compromise through unauthorized access to biometric data and artificial regeneration of such data (also termed spoofing), all of which can lead to identity theft [100]. These limitations of conventional biometrics have bolstered the investigation of novel biometric traits. Recently, novel biometric traits based on bio-signals, such as the electrocardiogram (ECG), electroencephalogram (EEG), and electromyogram (EMG) have been shown to be more resilient to spoofing than the conventional biometric traits [25]. For example, it could be relatively easy to take pictures and record videos discreetly and use printed photos to deceive a facial recognition system. On the contrary, obtaining biosignals would require more coordination and attaching specialized sensors to unwilling subjects, which can be used as a liveness detection ability [25].

Among the biosignals, surface EMG signal, which measures the electrical manifestation of skeletal muscle contractions, has been traditionally used in gesture recognition-based research, specifically for prosthetic control. Extensive investigation demonstrated EMG-based gesture control has poor cross-user transference performance [10]. In fact, a calibration-free EMG-based gesture recognition system that does not need new user training has been an elusive goal in myoelectric control literature. Such difficulty suggests that inherent individual differences exist in surface EMG signals. And this is precisely a biometrics trait. The surface EMG has been recently utilized in biometric applications as an accurate biometric trait [3, 30, 49, 74, 127]. Recent findings have suggested that EMG signals from the forearm skin surface, obtained while performing different hand gestures resulted in remarkably accurate biometric performance and

are resilient to spoofing [44]. The biometric capabilities of EMG signals have enabled their use in two distinct modes: authentication (one-vs-one), and, identification (one-vs-many matching).

In contrast with other physiological signals such as electrocardiogram (ECG) and electroencephalogram (EEG), a unique property of EMG is that it allows user-defined and distinguishable patterns by performing different hand gestures, enabling individuals to set personalized passwords comprising of multiple gestures. This unique property facilitates a ‘dual-security’, i.e., a system comprising both biometric-level and password-level security. A multi-code framework for a combination of gestures (codes) could be developed to facilitate such dual-security systems [106]. However, many of the current studies are limited by small database size (#participants < 25) and limited with data from one session/day [108]. In a multi-session scenario spread across days, non-stationary factors including electrode shifts, sweaty/dry skin, and physical conditions (rest vs. exercising) can affect the accuracy and consistency of gesture recognition based on EMG signals, challenging the performance of a biometric system. To create a robust biometric system invariant to these factors, it is important to 1) differentiate the genuine user from all other users (uniqueness) and 2) retain the biometric performance over multiple sessions and across multiple days (robustness).

Additionally, multiple studies use traditional feature extraction methods, which are not personalized or custom-trainable in nature for biometric applications. With the advent of deep learning and more powerful computational resources, it is possible to utilize models such as convolutional neural networks (CNNs) for learning features relevant to the true users only, and hence, resulting in higher accuracy. Some studies have utilized convolutional feature engineering for both authentication and identification applications [44, 62, 65, 82, 121]. However, these feature engineering studies not only lacked a comparison to the traditional feature engineering methods but also the investigation of different architecture types for utilizing spatial information from EMG channels.

In order to utilize EMG as a commercial biometric trait, it is necessary to achieve a secure EMG biometric system that 1) utilizes a multi-code framework for facilitating dual security, 2) is robust to multi-day use, and 3) can be custom trained (e.g., using convolutional networks). This will not only help achieve a secure biometric system but also address the current biometric limitations. Through this thesis, the above goals are addressed and the novel contributions are enumerated as follows.

1. Determined EMG system parameters, such as the feature extraction methods and the number of recording channels, for optimal biometric performances.
2. Developed a multi-code EMG biometric framework by comparing fusion strategies and determining the optimal codelength of such a system.

3. Recorded a large-scale EMG data acquisition involving 43 participants and three recording days over a period of 28 days, and was made open-access for researchers.
4. Investigated the multi-day performance of EMG biometrics using a multi-code framework for biometric applications.
5. Employed convolutional feature engineering on EMG signals for improving multi-day biometric performance. As compared to the traditional methods, the proposed feature engineering resulted in a significantly higher biometric performance.

The following diagram (Figure 1.1) depicts the connection between the different studies and their respective contributions.

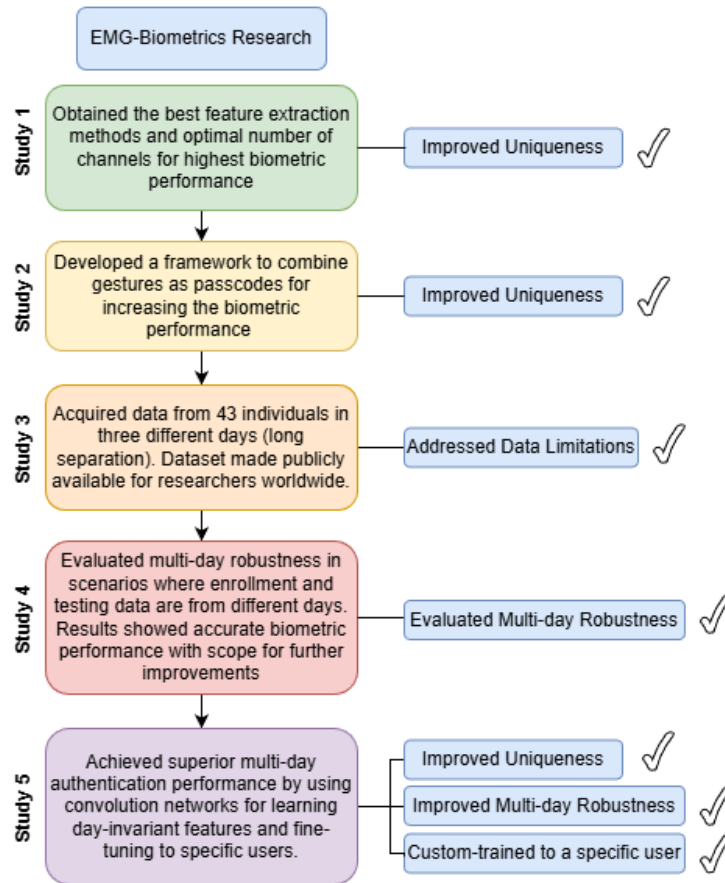


Figure 1.1: Flow diagram of the research studies

1.1 Thesis Organization

The outline of the thesis is listed here. The current conventional and novel biometrics are discussed in Chapter 2.1. EMG as a physiological signal, its system parameters, and its application in biometrics are discussed in Chapters 2.2 and 2.3. Fusion strategies for multimodal biometrics and a multi-code EMG framework are discussed in Chapter 2.4. In Chapter 2.5, Multi-day EMG Biometric performance is discussed. Previous research on convolutional feature engineering for EMG-based applications is discussed in Chapter 2.7.

Chapter 3, published as a stand-alone manuscript, investigates the optimal feature extraction and channel numbers for accurate EMG-based biometric performance. Chapter 4, published as a stand-alone published manuscript investigates the development of a multi-code framework for improved biometric performance as compared to the single-code setup. Chapter 5, published as a stand-alone published manuscript is a data descriptor paper that explains the data collection methodology as well the open-source accessibility of a multi-day and large-sample EMG Dataset. Chapter 6, published as a stand-alone published manuscript investigates the multi-day scenarios of biometric testing. Chapter 7, containing unpublished materials investigates the convolutional feature engineering for robust biometric performance and compares it with the traditional methods. The last chapter, Chapter 8, summarizes the key conclusions of this research and provides direction for future work.

Chapter 2

Literature Review

2.1 Biometrics

Biometrics are biological information or characteristics that can be used to identify an individual or verify an individual's identity. Biometrics (B) is one of the three traditional modes of a secure system [90]. The other two modes are possession, for example, the physical possessions of keys, IDs, and passports, while the other mode is knowledge (K), i.e., information known only to the true user such as a PIN or password. The use of biometrics in modern society is becoming increasingly important with the widespread application of ubiquitous smart and wearable devices and the associated risks of unauthorized access to personal information. The most-used biometric traits (e.g., fingerprint, iris, and facial recognition) are already embedded in our daily life through smartphones and laptops [143]. [Figure 2.1](#) shows a framework of an authentication system utilizing forearm EMG signals.

Ideally, a system with both knowledge and biometrics would be preferred to maximize security, as either mode has inherent drawbacks. The knowledge-based codes are liable to hacking and leakage. One of the advantages of the knowledge-based mode is that the code (password or PIN) can always be regenerated. On the other hand, common biometric traits, such as fingerprint, facial information, and iris scans, cannot be changed once they are compromised, posing a lifetime risk for the genuine user [100]. As such, there are many ways to compromise these traits by artificially generated spoofs without the knowledge of the legitimate user, leading to identity theft [100]. Therefore, a system with a dual security, i.e., simultaneous knowledge-based and biometric information, is highly desirable. Certain behavioral traits such as gait, voice, keystroke dynamics can be customized based on the user's knowledge, thus facilitating the dual-mode security. However, their low biometric performance have led to reduced acceptance of these novel

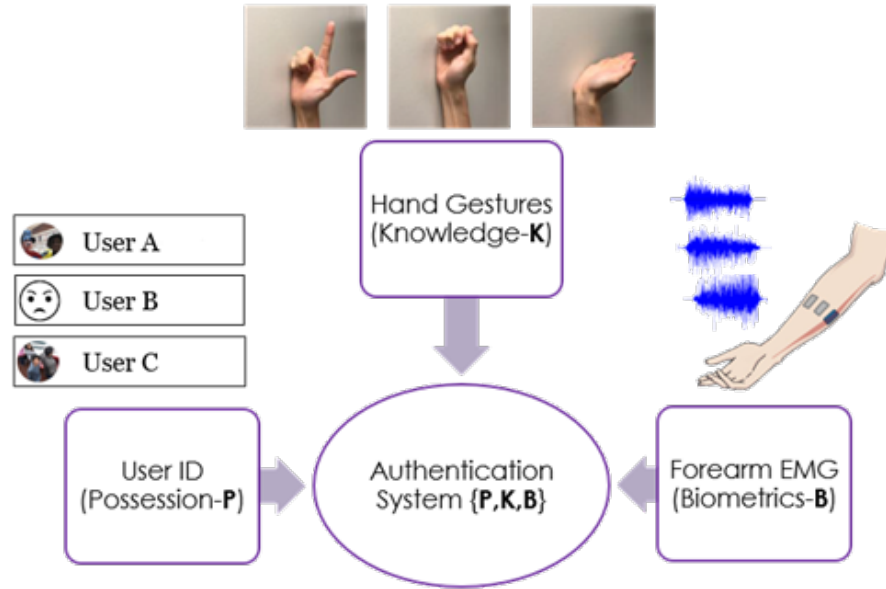


Figure 2.1: The framework of an EMG-based authentication system showing the three components: Possession, Knowledge, and Biometrics

biometric traits. Such low performances might be caused by changes in the immediate environment such as but not limited to physical health of individual, faulty sensing mechanism and noisy data-acquisition. Table 2.1 provides a list of the conventional and novel biometric traits and a comparison of their features.

A biometric system has two modes: 1) authentication and 2) identification. Authentication is the application of accepting or rejecting the claim of a user by comparing his/her biometric traits to that of the user's enrollment data stored in the system. Identification is the application of finding the closest match of presented biometric trait(s) among all the enrollment data stored with the system. In general, authentication applications have a higher requirement for the uniqueness of biometric traits than identification applications.

Recently, various biosignals or physiological signals such as electroencephalogram (EEG), electrocardiogram (ECG), and electromyogram (EMG) from the brain, heart, and the muscle, respectively, have been used as biometric traits [11, 41, 44, 100]. These biosignals are suggested as an auxiliary biometric trait because they have two important disadvantages compared to well-developed and established biometric traits. First of all, the performance accuracy is considerably lower [25]. Secondly, they are less robust against conditions such as illness, physical and mental activities, and the time interval between the training and the testing session [1, 79]. As such, re-

Table 2.1: Comparison of current biometric traits [25]

Biometric	Spoofing	Customizable	Environment factors	Accuracy
Fingerprint	Easy	No	Robust	High
Iris	Easy	No	Sensitive	High
Face	Easy	No	Sensitive	High
Voice	Easy	Yes	Sensitive	Low
Gait	Hard	Yes	Sensitive	Low
Keystroke Dynamics	Hard	Yes	Robust	Low
Biosignals				
EEG	Hard	No	Sensitive	Low
ECG	Hard	No	Robust	Low
EMG	Hard	Yes	Sensitive	Acceptable

search efforts have been undertaken to improve the accuracy and robustness of biometric systems based on biosignals. Although multiple studies have investigated EEG and ECG as biometric traits, there have been limited studies that use surface EMG from the forearm and hand muscles while performing hand gestures [44, 62, 65, 82]. The following section discusses more about the EMG signals and its applications.

2.2 Surface Electromyogram (EMG)

Electromyogram (EMG) signals are a graphical representation of electrical activity occurring in muscles during contraction [102]. The contraction of muscles generates a force that acts upon the skeletons which leads to body movements. Muscles comprise of fibers that shorten when activated by depolarization of cell membrane by nerve impulses. The muscle fibers are controlled by the central nervous system (CNS) and by the action of motor neurons. All of the muscle fibers innervated by one motor neuron form a motor unit (MU) and these muscle fibers always contract simultaneously. For activation of a motor unit, a local depolarization of the motor neuron membrane runs as a wave from the spinal cord to the muscles via the axons. This depolarization wave is called the action potential. A motor unit action potential (MUAP) propagates in both the directions of muscle fibers, activating them, and hence causing contraction of all fibers associated with the MU. The tissues overlying the muscles undergo volume conduction of the EMG signals, the individual MUAPs superimpose and form the surface EMG which could be obtained from the skin by using electrodes.

The amplitude of the EMG signal rises during a contraction and is of the order of few mil-

livolts. The frequency of EMG lies in the range of 0-450 Hz. The electrodes often used for research and clinical purposes are Ag/AgCl electrodes paired with a conductive gel. The gel reduces skin electrode impedance mismatch and thus prevents resulting half-cell potentials, which can result in a noisy signal. The surface EMG can be directly utilized as a monopolar signal (data recorded from one specific location), or as a bipolar or single-differential configuration (differential data recorded from two adjacent locations). The single-differential configuration contains spatial information regarding the conduction of EMG through the muscle length. Both the configurations have been utilized for EMG based applications. Figure 2.2 shows a graphical representation of EMG signals being collected from forearm.

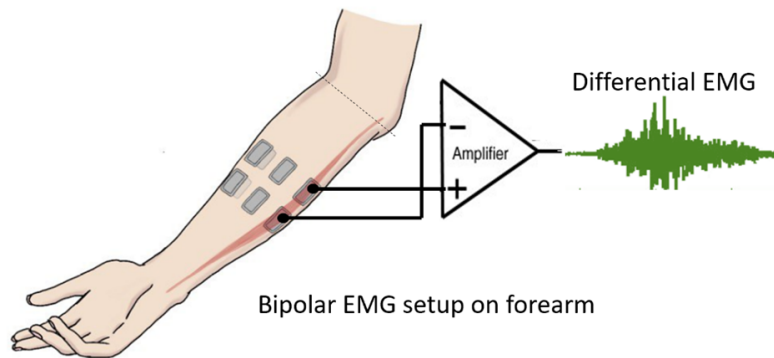


Figure 2.2: EMG signals collected from electrodes placed on the forearm. There are two types of signal acquisition: monopolar and single-differential EMG

Due its close correspondence with the muscle-force relationship, EMG can be used as a voluntary control signal for devices like myoelectric prosthesis [93, 110]. EMG has been predominantly used for accurate hand gesture recognition [28, 46, 123] due to its characteristic property that different movements result in distinctive EMG patterns. However, while training with multiple users, individual differences were observed [73, 89] which motivated the biometric studies based on EMG, as described in the next section.

2.2.1 EMG System Parameters

The performance of EMG-based applications can vary due to many factors. Specifically, the signal processing methods and the number of electrodes play a significant role in the accuracy of such applications.

Feature Extraction Methods

The three main categories of EMG feature extraction are: time domain [52], frequency domain [38, 47, 133], and time-frequency domain [133, 134, 138]. Additionally, there are non-linear feature extraction methods [53] that adhere to the non-stationary and complex nature of the EMG signals. However, it has been shown that for a short time contraction, the EMG signals satisfy the stationary assumptions [6]. Therefore, for a short duration, isometric contractions suitable for biometric applications [64, 65], the time domain and frequency domain features are ideal for feature extraction as their mean and variances have minimal variations [95]. These include Hudgin's time-domain (TD) feature set [52] and two frequency domain feature sets: Frequency division technique (FDT) [47, 48] and Autoregressive coefficients (AR) [47, 51]. Time Domain (TD) features are computed using the signal amplitude values. The computed features, namely mean absolute value (MAV), slope sign changes (SSC), zero crossing (ZC), and waveform length (WL), quantify EMG waveform amplitude, frequency, and duration as distinguishing characteristics. Due to their relatively lower computational cost, TD features have been preferred in multitudes of pattern recognition-based research [123].

The frequency-domain category of features is based on the principle that the EMG spectrum varies with limb movements and has shown robustness in real-time myoelectric control of prosthetic functions [109, 137]. The AR feature set includes the coefficients of the n th order autoregressive model of the EMG signal [15, 38, 51] and has shown to be stable over varying muscle force levels and shifts in electrode location [136]. To the best of our knowledge, there is no existing research on the comparison of feature extraction methods for EMG based biometrics. Therefore, it is crucial to determine the best feature-set that not only has good gesture recognition accuracy but also considers individual differences between users.

Number of Channels

In addition to the feature extraction method, the number of channels is also an important factor affecting the performance of EMG based gesture recognition. It has been shown that increasing the number of channels, as well as the use of HD grids, provides a more accurate gesture recognition [26, 113]. However, space limitations, power consumption, and viability limitations such as setup-time and connection to data-acquisition devices, specifically in a practical scenario, hinder the use of a large number of EMG channels. In this regard, previous studies have analyzed the effect of channel reduction on gesture recognition performance [50, 150]. It was shown that four to six channels are sufficient for gesture recognition [150]. It has been shown that a uniform electrode positioning of five electrodes around the forearm, regardless of muscle anatomical location provided a high accuracy of gesture recognition [13]. Therefore, research

on channel reduction is important to assess performance in EMG based biometric systems, thus facilitating the development of easy-to-use wearable sensors

2.3 EMG-Biometrics

An interesting finding in EMG-based gesture control research is that when EMG models are trained by data from one individual, and tested on another individual, significant performance degradation occurred [124]. Many studies have investigated different approaches to realize a cross-subject or an user-independent EMG gesture recognition system, without satisfactory outcome [73, 89]. Cross-subject or user-independent EMG systems are still considered an open problem. Recently, advanced deep learning techniques have been explored to address this problem with limited success [21]. The difficulties in realizing such a one-size-fits-all EMG gesture recognition is that the presence of individual differences in the EMG signals, indicating that EMG signals have great potential as a biometric trait.

A unique feature of EMG is that it allows an inherently dual-mode security: 1) Knowledge (K) level security, where the user can set customized hand gestures as ‘codes’, and 2) biometric (B) nature of EMG that can identify individuals. Previous studies on the topic implicitly or explicitly utilized this unique advantage of dual security over ECG and EEG signals [44]. As the concept of using EMG as an authentication or identification is relatively new, the existing literature is limited [44, 62, 65, 82, 121]. Initial studies focused on the classification of individuals using different pattern recognition algorithms and only reported the classification accuracy [125, 127, 145]. The lack of comparisons to conventional traits, such as fingerprints, iris, and face, limits interpretation of results towards establishing EMG as a biometric trait.

One EMG-based study utilized the two applications of biometrics, authentication and identification, and reported an equal error rate (EER) of lower than 10% [44]. The study involved 24 participants who performed 16 hand gestures. The frequency division technique (FDT) was used for the feature extraction method [48]. Other studies used a high-density EMG (HD-EMG) setup where 64 channels were placed in a grid formation on the dorsal side of the hand [62, 65]. The feature extraction involved a combination of time-frequency-spatial domain features as well as HD-EMG decomposition-based features. A recent study utilized only one electrode placed on the flexor digitorum superficialis muscle while the participants performed a phone unlock pattern with contractions of their fingers [82]. The feature extraction technique and the number of channels can change the system performance; therefore, it is important to investigate their effects on the biometric performance and to determine an optimal processing configuration for using EMG based biometric applications.

The concept of using a gesture sequence as a neuromuscular password was introduced in a previous study [62]. A different authentication study allowed users to choose three out of four gestures as their authentication code [121]. An identification study with a similar setup as in [121] presented EMG biometrics as a variant across multiple codelengths (or the number of gestures) [62]. It was shown that with increasing codelength, the biometric performance increased. However, all these studies used EMG signals of the entire sequences in the algorithm training, limiting the flexibility of the biometric system. Additionally, when a user changes their code sequence, a new training process is required. With knowledge-level security, the user should be able to customize their code sequence as a combination of multiple codes as desired. However, the framework of a multicode EMG biometric system, in which the maximum potential of dual security is realized, has not been systematically examined for authentication or identification systems.

Some of the EMG-based studies have reported a high biometric authentication [44, 61, 64, 83, 105, 121, 145] and identification performance [9, 30, 49, 60, 62, 75, 87, 101]. Previous studies on EMG-Biometrics included varying sample-sizes, recording days, hand gestures, and number of recording channels (shown in Table 2.2). The number of hand gestures in these studies ranged from 1 to 34. The number of EMG channels varied from as low as one channel to as high as 256 channels. While most of the studies had fewer than 25 subjects, only three studies included larger numbers (> 40) of subjects [29, 121, 126], more appropriate in the biometric context. However, most studies in the literature were limited to data acquired within only one session or one day. Table 2.2 lists recent studies on EMG-based biometrics. Additional details including the number of sessions, participants, EMG channels, and gestures, are also provided.

Table 2.2: List of recent studies on EMG based biometric authentication and identification

Reference	Application	sessions	participants	EMG channels	Gestures	Error Rate
Jiang (2020)	Identification	2	22	64	34	Accuracy - 99.5%
Pradhan (2020)	Authentication Identification	1	24	4	16	Authentication-4% Identification-3%
Li (2020)	Authentication	1	10	1	2	Half of total error-10.01%
Hu (2021)	Identification	1	40	8	40	Rank-1 accuracy - 74%
Pleva (2022)	Identification	4	5	8	75	-
Jiang (2020)	Identification	2	22	64	34	85.8%
Jiang (2022)	Identification	2	20	256	34	99.85%

Fan (2021)	Identification	2	22	64	34	Accuracy -87.23% EER-4.66%
Kim (2022)	Identification	3	200	2	12	Accuracy -74.19%
Lu (2020)	Identification	1	21	8	1	Accuracy - 99.2%
Raurale (2021)	Identification	4	5	8	8	Accuracy - 92%
Fan (2021)	Authentication	1	53	8	1	Acceptance 91.81% FAR - 7.43%
Shin (2021)	Authentication	1	50	2	1	Accuracy - 87.1%
Said (2020)	Authentication	1	56	8	4	FAR 0.2% FRR 2.9%
Jiang (2020)	Authentication	2	22	64	34	14.96%
He (2020)	Authentication Identification	1	24	8	16	Authentication-3.5% Identification-9.7%
Jiang (2021)	Authentication	2	20	256	34	-
Pradhan (2021)	Authentication Identification	1	24	4	16	Authentication-0.6% Identification-1.14%
Li (2022)	Authentication (two datasets)	1	18 & 27	128 & 10	8	EER (HD) - 0.023 & EER (SD) - 0.061
Yamaba (2018)	Identification	4	8	8	3	95%
Qiu (2023)	Authentication (two datasets)	2 & 2	24 & 20	16 & 256	12 & 12	EER - 0.0018
Hu (2024)	Identification (two datasets)	1 & 2	40 & 20	12 & 256	40 & 16	96.9% & 80%

2.4 Multi-code framework

A biometric system can be of a single-code or multicode nature [120]. A single-code biometric system utilizes single source information (data sample) pertaining to a specific user and comprises five modules [68]: 1) a sensor/data module that collects the biometric data; 2) a feature extraction module that processes biometric data to extract characteristic features of individual performance; 3) a matcher module that compares the extracted feature vector to the enrollment data stored in the system database and generates a score; 4) a ranking module that ranks the

scores of all the users (typically used in an identification system); and 5) a decision module that makes a decision based on the rank of the user in the biometric system. A schematic diagram of a typical single-code biometric system is shown in Figure 2.3 (A). In an authentication system, the ranking module is absent, and the decision is made based on the scores (i.e., access is granted or rejected). In an identification system, the usual practice is that the K identities with the highest K scores are chosen. If the claimed identity is one of those K identities, an authorization/rejection of the claim is rendered. This procedure is called rank- K identification; in an ideal scenario, the genuine user is the top-ranked user in the biometric system. However, to reduce possible false rejection, a lower rank ($K > 1$) is also commonly used.

In contrast to single-code biometric systems, multicode biometric systems utilize multiple source information that can be at any of the above-mentioned stages to identify or verify a user. Currently, most biometric systems are of a multicode nature that provides a higher level of security than single-code biometric systems [120]. Fusion is an important underlying mechanism of a multicode biometric system. Fusion between codes can occur at five different levels: sensor-, feature-, score-, rank-, or decision level. Based on the application and biometric traits, the most appropriate fusion rule must be chosen to achieve the highest-possible security level [91].

In a multicode EMG biometric system the user performs multiple gestures; specifically for an authentication application, such knowledge-based information would significantly enhance the security level of the system, as the user has to remember the correct codes (gestures) and their correct order. In a multicode system, each code will be processed by a pipeline with the same structure as a single-code system (see Figure 2.3 (B)); biometric information from different codes or gestures can be integrated or fused at different levels of the processing pipeline. As mentioned earlier, the previous multicode EMG studies used signals and features from all sequences that correspond to the sensor- and feature-level fusion [62, 121]. As these fusion strategies limit the code-customization flexibility of biometric systems, they are not considered in this study. Three post-matching fusion strategies are proposed for the multicode EMG biometric system: Score-, rank-, and decision level as explained below and illustrated in Figure 2.3 (B). It should be noted that the rank-level fusion is applicable only for identification applications.

Score-level fusion: One of the most common score-level fusion methods is the simple summation of the scores of multiple matcher modules [119]. It was expected that these scores from multiple codes would vary in their respective distributions, therefore, normalization was performed prior to the summation step to address these variations [67, 68].

Rank-level fusion: In a rank- K identification, the output of each code is a vector, i.e., $r = [r_1, r_2 \dots r_j \dots, r_J]$, where r_j is the rank of the j^{th} enrolled user to be the user currently presenting biometric information. In a single-code case, access is granted to K users with the highest probabilities. In the multi-code case, the rank-level fusion combines the rank outputs of

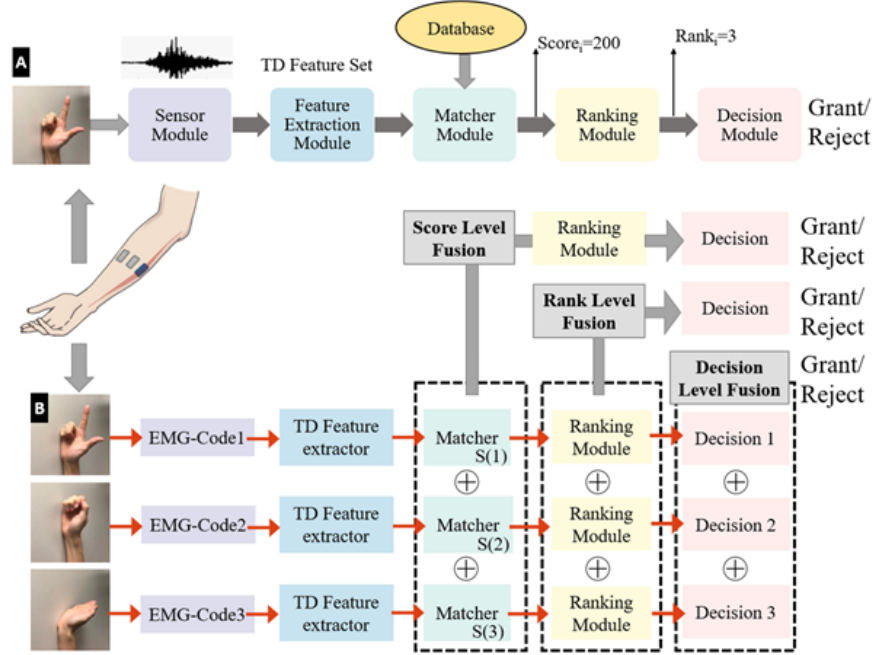


Figure 2.3: A) Single-code EMG Biometric system. The processed input is matched with the template(s) in the database and a score is generated for every matching operation. B) Multicode EMG biometric system combines individual codes using a post-matching fusion at one of the three levels—score, rank, and decision. For both systems, the ranking module is present only for identification, whereas the authentication application bypasses the ranking module, and a decision is made based on the score threshold.

all codes using two methods, the Borda count and logistic regression methods [91], to provide an enhanced level of precision.

Decision-level fusion: In a multi-code biometric system, the simplest decision-level fusion method is obtained by using the logical AND rule on the outcomes of individual codes (Decision-AND). The output of the Decision-AND is a correct match when all the codes agree that the user presenting biometrics matches the template. Another method is to use majority voting.

2.5 Multi-day Robustness of EMG-Biometrics

One of the most important features of a viable biometric trait is its longitudinal robustness across multi-session and multi-day. However, most studies in the literature were limited to data acquired within only one session or one day. Two studies with a small subject pool (<22), utilized data from a two-day protocol for biometric analysis [61, 62]. Only one study had five subjects with a four-day data collection protocol [117]. It has been established in the EMG processing literature, that in a multi-session protocol spreading across days, non-stationary factors including electrode shifts, sweat, dry skin, and physical conditions will affect the accuracy and consistency of the EMG processing system [5]. Therefore, the multi-day performance with a sufficiently large subject pool is a crucial step for validating the effectiveness of EMG as a biometric trait.

2.5.1 Multi-day Biometric Performance

While some multi-day studies analyzed EMG-Biometrics (shown in Table 2.1), a systematic comparison is challenging due to the differences in experimental protocol, feature extraction strategies, biometric analysis, number of gestures in the code (codelength) and fusion strategies used. One study investigated a password-based EMG biometrics of codelength eight and reported an *EER* of 0.012 when the code is not known to the testing users (normal test) and an *EER* of 0.1496 [64] when the code is known to the testing users (leaked test). A different study involving a similar multi-code password of codelength 12 reported a normal-test *EER* of 0.0013 and the leaked-test *EER* of 0.0273 [61]. Another study incorporated a multi-day investigation, however, different codelengths were not investigated. They reported an average *EER* of 0.1634 for eight individual gestures [117].

In an identification study, multiple protocols, codelengths, and electrode layouts for improving biometric performance were investigated [60]. For a fixed gesture sequence protocol and a codelength of four, a mean *rank-1 accuracy* of 86.1% was reported. The sample size in the study was 22 and the number of electrode channels was 256. Another study with a similar setup of 22 users and 256 channels reported a *rank-1 accuracy* of 85.8%, however, the gesture task of each participant was different, and they reported a decrease in accuracy ($<50\%$) with more than 30 inactive electrodes [62]. The high-density electrode setup (256 channels) can have multiple limitations such as the practical viability of acquisition-setup, inactive channels, and model training time. Another study with a similar setup (22 gestures and 256 channels) and a codelength of eight used a deep learning model and reported a *rank-1 accuracy* of 87.2% and a *rank-5 accuracy* of 91.2% [30]. Two other studies involved multi-day testing data from 5 subjects and reported an identification accuracy of 74.1% and 93% [75, 117].

While previous findings establish the potential of EMG-Biometrics, further research is warranted to improve the multi-day authentication and identification performance. For a comparable performance with the conventional biometrics (fingerprints, facial recognition, and iris scan), the analysis should include a larger subject pool and multiple days of EMG recording. This will bolster further research on the biometric ability and robustness of EMG signals.

2.5.2 Multit-day Forearm and Wrist Electromyogram Datasets

Some open-access databases of multi-day EMG recordings of forearm muscles are publicly available [2, 17, 59, 69, 139]. Two databases with large subject pool (>40) involved two days of data collection [69, 139]. While one dataset employed six channels [139], some others utilized a high-density (HD) EMG setup [2, 17, 59]. Four studies involved more than two data collection days, but the number of subjects was smaller (<11) [2, 23, 31, 36]. Only one study with three days of data recording had a sample size of 20 subjects, with only signals from the forearm [23]. Table 2.3 lists the multi-day and open-access datasets that are currently available for further research.

Table 2.3: Multi-day forearm and Wrist EMG Datasets

Database	Subjects	Days	Gestures	Channels
CapgMyo [2]	10	2	8	128 (HD)
CSL-HDEMG [17]	5	5	27	192 (HD)
putEMG [69]	44	2	8	24
DS-DataSet [139]	42	2	7	6
Hyser [59]	20	2	34	256 (HD)
Megane Pro [36]	10	5	7	14
ISRMyo-I [31]	6	10	10	16
Ninapro DB6 [66]	10	5	7	14
Long-Term sEMG Dataset [23]	20	3	11	10

To explore the robustness and accuracy of gesture-recognition and biometrics, it is imperative to have a database with larger subject pool sizes, recorded across multiple days, and comprising of numerous gestures.

2.5.3 Wrist as a preferred location for EMG

Most studies in the literature used EMG data acquired from the forearm, which is not convenient for consumer-based applications, such as biometrics. There has been a recent change in focus from the forearm to the wrist, specifically for gesture recognition applications for more general-consumer use [8]. The wrist presents a more feasible and attractive position for biometrics applications because wrist-worn devices are well-established and ubiquitous. For this purpose, the current study utilizes EMG signals collected from the wrist while performing hand gestures. A wrist electrode setup will facilitate the research and development of industry-grade wearable wristbands, which have previously been explored for gesture recognition [58] and biometric authentication applications [126, 127].

2.6 Convolutional feature engineering using EMG

With the recent advent of deep learning methods in biomedical applications, these methods have been adopted for processing EMG signals for hand gesture recognition [54, 144]. Conventional hand gesture recognition (HGR) utilized pattern recognition methods where feature extraction was a crucial step followed by a classification task where a certain gesture class is predicted. With deep learning, feature engineering is combined with classification tasks, facilitating end-to-end learning. In one study, a convolutional neural network (CNN) was used on instantaneous images of EMG (2D or spatial representation of the EMG channels for each timepoint) [35], resulting in better classification performance than traditional methods. In a recent study, CNN outperformed other time series-based deep learning methods such as recurrent neural network (RNN), long short-term memory (LSTM), gate recurrent unit (GRU), and a transformer network [45]. This suggests that although the EMG is a time-series data, the HGR application lacks a seasonal component, or the natural frequency an event repeats itself, which are crucial for time-series models. Hence, the CNN is better suited for EMG applications than the other deep learning methods. Figure 2.4 shows the structure employed in the previous study reported in [45].

Based on the input EMG data (signal processing window size and the number of channels), CNN can have different architecture types for utilizing spatial information from EMG channels. Similar to two adjacent pixels in an image, two EMG electrodes placed close to each other share ample similar information. Convolutional filters can accommodate neighboring channels/pixels and utilize spatial information to the benefit of the classification task.

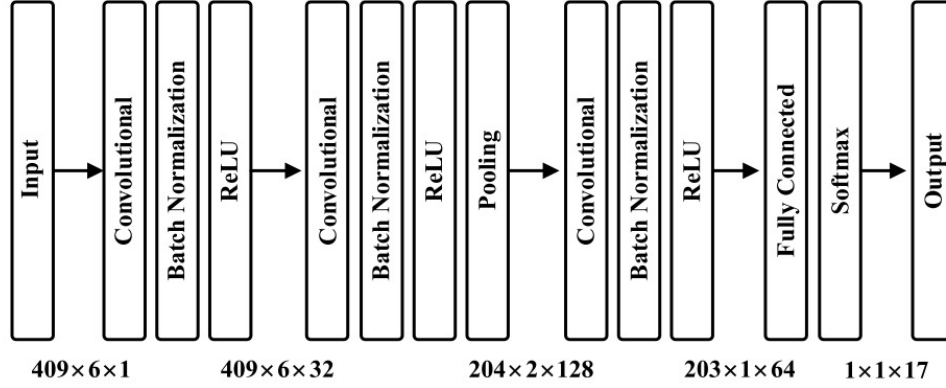


Figure 2.4: The structure of the convolutional neural network used in a gesture recognition study [45]. The study used three convolutional blocks and one fully connected layer block to match the input (409 frames x 6 channels) and output (17 classes)

2.6.1 CNN in EMG-Biometrics

As the biometric application of EMG signal is a more recent investigation, few studies have investigated CNNs. Convolutional feature engineering has been employed on forearm and wrist EMG signals in biometric applications [3, 30, 49, 74, 127]. In a cross-day study, a three-layer convolution network utilized wrist EMG for authentication. The study found comparable performance between low- (16 channels) and high-density (256 channels) systems, which facilitated the development of a real-time application [114]. In another study, a deep anomaly detection strategy was used for the authentication application where EMG feature maps were utilized in a generative network with convolution layers [83]. Similarly, a Siamese network with three convolutional layers used EMG signals from a phone-holding action to unlock a smartphone for the true user [29]. To improve the robustness of electrode shift, additional training data is recorded from rotated electrode locations. In another study, a three-layer convolutional network was used for authentication and identification applications over four different days [127]. Additionally, the network employed two different output heads allowing simultaneous predictions of gestures and user authentication. Furthermore, some studies also utilized convolutional feature engineering for EMG-based personal identification [30, 49, 74, 87]. While one study suggested a cancelable property of EMG signals [30], *i.e.*, if one code is leaked, an alternative code can be used, another study explored a gesture-invariance property where identification was invariant to different gestures [49].

Although previous studies utilized similar CNN architectures, the weights are trained to minimize custom loss functions, thus facilitating the exploration of different biometric prop-

erties. Various signal processing approaches have been explored, such as time domain feature maps [83], 2D spectrograms [30, 49, 74], and time-frequency maps [87]. Other studies have utilized bandpass filtering signals as 1D data over multiple channels [29, 83, 127]. This is advantageous as it reduces computation time and avoids the loss of information. The previous CNN studies not only lacked a comparison to the traditional feature engineering methods, but also the investigation of different architecture types for utilizing spatial information from EMG channels.

Therefore convolutional feature engineering is recommended to process raw EMG signals in an accurate biometric application. Specifically for the authentication mode, a more daily-used application, CNNs can be custom-trained for each individual resulting in a secure gesture-based biometric password that is robust to multi-day applications.

2.7 Research Objectives and Hypothesis

As EMG signals are a novel biometric trait, as compared to conventional traits such as fingerprint and facial recognition, research on optimal parameters, multi-code framework, continuous authentication framework, large dataset performance, and multi-day robustness is warranted. A series of studies are proposed and their objectives are listed below:

1. **Optimal Parameters:** The purpose of the study was 1) to compare different EMG feature extraction techniques and their combinations on the authentication performance, 2) to study the effect of channel reduction on the authentication performance, 3) to determine the optimal configuration (feature extraction method and the number of channels) for accurate authentication performance using EMG based biometrics.
 - H1a:** It was hypothesized that the biometric performance of time-domain (TD) feature extraction will be significantly better than frequency-domain techniques .
 - H1b:** It was hypothesized that the biometric performance will increase while increasing the number of channels.
2. **Multi-code Framework:** The aim of the study was 1) to compare the biometric performances of the three fusion strategies (score-level, rank-level and decision-level) and 2) to determine the optimal parameter (codelength and rank-K) values for the three fusion methods.
 - H2a:** It was hypothesized that the biometric performance of multi-code EMG-Biometrics framework will be significantly higher than the single-code framework.
 - H2b:** It was hypothesized that there will be a difference between the biometric performance of different fusion strategies.

H2c: It was hypothesized that there will be a difference between the biometric performance while increasing the codelength and rank (identification mode only).

3. **Multi-Day and Large-Sample Dataset:** The aim of the study was to acquire a high-quality EMG dataset with the following properties: 1) large subject-pool sample size ($\# \text{ subjects} > 40$) and 2) multiple recording days ($\# \text{ days} \geq 3$) 3) multiple hand/wrist gestures ($\# \text{ gestures} \geq 16$) and 4) multiple repetitions of each gesture ($\# \text{ repetitions} \geq 5$)

4. **Multi-day analysis:** The aim of the study was to analyze the multi-day biometric performance of wrist and forearm EMG using a multi-code framework. A multi-day analysis involving training data and testing data from different days was employed to test the robustness of the EMG-Biometrics in practical scenarios.

H4: It was hypothesized that the biometric performance of multi-code EMG framework will be significantly higher than the single-code framework in various multi-day analysis scenarios.

5. **Convolutional Feature Engineering:** The aim of the study is to compare the cross-day authentication performance of convolutional and traditional feature extraction methods, and 2) to investigate the effect of different training paradigms and network architecture on authentication performance.

H5a: It was hypothesized that there will be a difference between the cross-day authentication performance of the traditional feature engineering and the convolutional feature engineering.

H5b: It was hypothesized that there will be a difference between the cross-day authentication performance of different convolutional architecture types and training paradigms.

Chapter 3

Performance Optimization of Surface Electromyography Based Biometric Sensing System for Both Authentication and Identification

This chapter was published in *the IEEE Sensors Journal*, vol 21 as Pradhan, A., He, J., & Jiang, N., “Performance optimization of surface electromyography based biometric sensing system for both verification and identification”, article 19, pages 21718-21729, 2021 [105]. This chapter investigates the EMG system parameters such as feature extraction methods and number of channels to determine the optimal configuration for accurate biometric performance. As a note to the reader: the term ‘verification’ in the published document has been interchanged with ‘authentication’ to align with the remaining parts of the thesis.

3.1 Chapter Introduction

Biometrics-based security systems are critical and integral components of modern societies, can be found in personal electronic devices, law enforcement, and airport security [143]. Modern security systems usually operate on two modes: knowledge (of a unique code) and biometrics, respectively [7]. The knowledge-based security mode usually takes in the form of a unique code, known only to the genuine user. Passwords and PIN (personal identification number) are the most common choices of this security mode. On the other hand, the biometrics-based security mode

utilizes traits and characteristics that are uniquely specific to an individual, such as fingerprints or iris. These two modes can be incorporated separately but also be combined to provide a higher level of security. In the biometric mode, there are two applications: 1) authentication and 2) identification. Authentication is the application of accepting or rejecting the claim of a user by comparing his/her biometric traits to that of the user's enrollment data stored in the system. Identification is the application of finding the closest match of a presented biometric trait(s) among all the enrollment data stored with the system. In general, identification applications have a higher requirement for the uniqueness of biometric traits than authentication applications. As such, a system with a dual-security mode i.e., comprising simultaneous knowledge-based and biometric information is highly desirable.

Recently, surface electromyography (EMG) emerged as a novel biometric trait. As the concept of using EMG as an authentication or identification method is relatively new, the existing literature is limited. Initial studies focused on the classification of individuals using different pattern recognition algorithms and only reported the classification accuracy [125, 127, 145]. This made the results difficult to establish EMG as a biometric trait by comparing it to current biometric traits, such as fingerprints, iris, and face. A different EMG-based study utilized the two applications of biometrics, authentication and identification, and reported an equal error rate (*EER*) of lower than 10% [44]. The study involved 24 participants who performed 16 hand gestures. The frequency division technique (FDT) was used for the feature extraction method [48]. Other studies used a High-density EMG (HD-EMG) setup where 64 channels were placed in a grid formation on the dorsal side of the hand [64, 65]. The feature extraction involved a combination of time-frequency-spatial domain features as well as HD-EMG decomposition-based features. A recent study utilized only one electrode placed on the flexor digitorum superficialis muscle while the participants performed a phone unlock pattern with contractions of their fingers [82]. The time-domain feature extraction method was employed on the EMG signals. All these above studies showed a low error rate ($\leq 15\%$) of average biometric performance, indicating EMG as a promising biometric trait. However, considering many parameters of EMG processing system, especially the feature extraction methods and the number of electrode channels [80, 98] can change the system performance, it is important to investigate their effects on the biometric performance and to determine an optimal processing configuration of using EMG-based biometric applications.

Therefore the purpose of the study was 1) to compare different EMG feature extraction techniques and their combinations on the biometric performance, 2) to study the effect of channel reduction on the biometric performance, 3) to determine the optimal configuration (feature extraction method and the number of channels) for accurate biometric performance using EMG biometrics.

3.2 Methods

3.2.1 Participants

Twenty-four healthy participants (13 males and 11 females, age: 19-30 years) were recruited for the study. The average forearm length (measured from the styloid process to the olecranon) was 25.4 ± 1.46 cm. Participants were informed that they may withdraw from the study at any time during the experiment session. The experiment protocol was in accordance with the Declaration of Helsinki and approved by the Office of Research Ethics of the University of Waterloo (ORE #: 22391). Additionally, data from 20 participants (age: 22-30), publicly available through the NinaPro DataBase DB7, was used for the comparison of results [76].

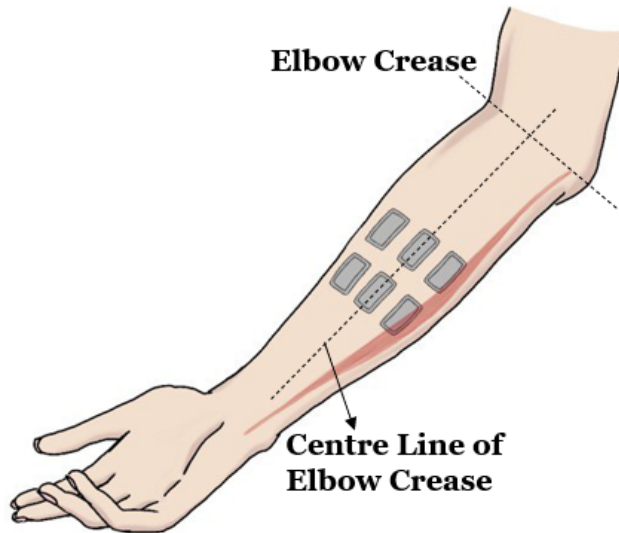


Figure 3.1: Positions of the sixteen monopolar electrodes on the forearm (dorsal view). In each ring, eight electrodes (grey rectangles on the forearm) are evenly placed, and one electrode is placed in the centerline of the elbow crease. The distance between the centers of the two rings was 2 cm, and the proximal ring was placed one third of the forearm length between the elbow crease and the upper ring.

3.2.2 Experimental Protocol

The participants were seated comfortably in a height-adjustable chair in an upright position with both of their upper limbs in a resting position and pointing towards the ground. Sixteen monopolar EMG electrodes (AM-N00S/E, Ambu, Denmark) were placed in the form of two rings, each consisting of eight electrodes equally spaced around the forearm and forming bipolar pairs (Figure 3.1). To make the electrode position more uniform across all participants, one of the electrodes in each ring was positioned in the centerline of the elbow crease. The center-to-center distance was 2 cm between the two rings, and the proximal ring was at a distance of one-third the forearm length from the elbow crease. A computer screen placed in front displayed visual instructions to help the participant perform the defined gestures. The EMG signals were acquired using a commercial amplifier (EMG USB2+, OT Bioelettronica, Italy). The sampling rate was 2048 Hz, and a bandpass setting of the device was set at 10 Hz and 500 Hz. Eight bipolar channels were created from data by taking the differential of each paired electrode from the two rings and used for subsequent processing.

The following sixteen hand and wrist gestures were included in the current study (Figure 3.2): lateral prehension (LP), thumb adduction (TA), thumb and little finger opposition (TLFO), thumb and index finger opposition (TIFO), thumb and little finger extension (TLFE), thumb and index finger extension (TIFE), index and middle finger extension (IMFE), little finger extension (LFE), index finger extension (IFE), thumb extension (TE), wrist flexion (WF), wrist extension (WE), forearm supination (FS), forearm pronation (FP), hand open (HO), hand close (HC). Each gesture was repeated seven times and five seconds of recording was performed each repetition. The dataset has been made publicly available. To avoid EMG signals from transitory movements, the participants were asked to start the contraction a little bit earlier than the beginning of the recording. A five-second rest period was provided between two consecutive contractions to avoid muscle fatigue. The order of the gestures was randomized for each participant. For the public dataset (NinaPro DB7), participants performed 17 isometric and isotonic, hand and finger gestures. Six repetitions of each gesture were performed and 5 seconds of data were recorded using a Delsys Trigno IM Wireless EMG system [76].

3.2.3 Signal Processing and Feature Extraction Methods

The bipolar surface EMG signals were first windowed with a window length of 200 ms and an overlap of 150 ms, i.e. 50 ms overlap between two consecutive windows. Each window was processed subsequently using three commonly used EMG feature extraction techniques: 1) Hudgin's Time Domain (TD) feature extraction [123], 2) Frequency Division Technique [11], and 3) the autoregression (AR) technique [89]. For the TD feature set, time-domain features

(mean absolute value, zero crossing, slope sign changes, and waveform length) were extracted from filtered data. The mathematical form of these features is provided below. For a signal channel of EMG signal x ,

$$\text{RMS} = \sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2} \quad (3.1)$$

$$\text{SSC} = \sum_{i=2}^{N-1} f(x_i - x_{i-1})(x_i - x_{i+1}) \quad (3.2)$$

$$f(x) = \begin{cases} 1, & \text{if } x \geq T_h, \\ 0, & \text{otherwise.} \end{cases}$$

$$\text{WL} = \sum_{i=1}^{N-1} |x_{i+1} - x_i| \quad (3.3)$$

$$\text{ZC} = \sum_{i=1}^N \text{sgn}(x_i \times x_{i+1}) \cap |x_i - x_{i+1}| \geq T_h \quad (3.4)$$

$$\text{sgn}(x) = \begin{cases} 1, & \text{if } x \geq T_h, \\ 0, & \text{otherwise.} \end{cases}$$

For FDT, the signals from each channel were divided into specific frequency sub-bands as given by [11]:

$$\text{FDT}_i = F \left[\sum_{j=1}^{n_i} |X(f_{i,j})| \right], i = 1, 2, \dots, L, \quad (3.5)$$

where $X(\cdot)$ denotes the magnitude of the FFT spectrum, and $F(\cdot)$ denotes a non-linear transformation (commonly a logarithmic operator) to obtain a smooth value for better classification results. In the current study, the whole EMG frequency band (20-450 Hz) is subdivided into six equal-width frequency bands: 20-92, 92-163, 163-235, 235-307, 307-378, and 378-450 Hz, consistent with prior studies [145] AR model is a prediction model defined as:

$$x_i = \sum_{p=1}^P a_p x_{i-p} + w_i \quad (3.6)$$

For the AR feature set, the 6th order autoregression coefficients (a_p) were used for it was widely adopted in previous studies [89]. Therefore, the three different feature vectors (TD, FDT, and AR) were analyzed separately.

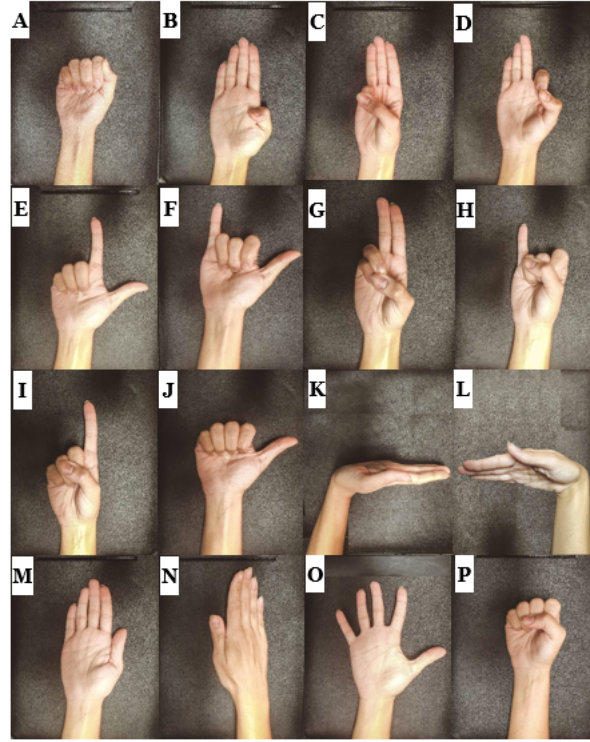


Figure 3.2: Sixteen gesture classes investigated in the study: (A) lateral prehension (LP), (B) thumb adduction (TA), (C) thumb and little finger opposition (TLFO), (D) thumb and index finger opposition (TIFO), (E) thumb and little finger extension (TLFE), (F) thumb and index finger extension (TIFE), (G) index and middle finger extension (IMFE), (H) little finger extension (LFE), (I) index finger extension (IFE), (J) thumb extension (TE), (K) wrist flexion (WF), (L) wrist extension (WE), (M) forearm supination (FS), (N) forearm pronation (FP), (O) hand open (HO), (P) hand close (HC).

3.2.4 Channel Reduction

A channel selection algorithm based on the sequential forward selection (SFS) method was implemented to examine system performance when only a subset of the available EMG channels is

used. The procedure of the SFS is described below. Two channel sets were defined: the applied channel set A (initially empty, i.e. $A = \phi, m = 0$) and the remaining electrode set R (initially full, i.e. $R = \{1, 2, 3 \dots, N\}, n = N$). The number m is the size of A , n is the size of R , and N is the total number of available channels (8 in the current study).

In each iteration, a union of A and one of the channels from R , i.e., r_k was used for feature extraction and subsequent biometric system evaluation. This step is repeated for all channels in R . Then, the channel that produces the minimum EER in the case of the authentication system and minimum $rank - k$ error ($k=1,5$) (alternatively termed $R1E$ and $R5E$) in the case of the identification system was chosen, respectively. These errors are discussed in the following sections. Thus, for the j^{th} iteration:

$$r_j = \operatorname{argmax}_{r_i} \left\{ EER_{A \cup \{r_i\}} : EER_{A \cup \{r_i\}} = f_{EER}(A \cup \{r_i\}) \right. \\ \left. \text{for } r = 1, 2, \dots, R(n) \right\} \quad (3.7)$$

$$r_j = \operatorname{argmax}_{r_i} \left\{ RkE_{A \cup \{r_i\}} : RkE_{A \cup \{r_i\}} = f_{RkE}(A \cup \{r_i\}) \right. \\ \left. k = 1, 5 \text{ \& for } r = 1, 2, \dots, R(n) \right\} \quad (3.8)$$

The applied set and remaining set are updated as:

$$A = A \cup \{r_j\}, \quad R = R \setminus \{r_j\}, \\ m = m + 1, \quad n = n - 1 \\ \text{for } j = 1, 2, \dots, N \quad (3.9)$$

The iterations are repeated till R is empty and A contains all N channels. The evaluation parameters for the authentication system (EER) and identification system ($R1E$ and $R5E$) in each iteration stage are reported for analysis. Also, the range namely r , which is the difference between the maximum and minimum error for every iteration is reported to analyze the distribution of error. Therefore, EER_r for authentication and $R1E_r, R5E_r$ for identification were reported, respectively. The error ranges are a measure of electrode location significance (for a particular iteration of channel selection, the higher the error range, the more significant the specific channel selected).

3.2.5 Performance Evaluation of the Authentication System

The performance analysis of the authentication system was performed using the detection error tradeoff curve (DEC), where false rejection rate (FRR) was plotted against false acceptance rate

(*FAR*) for various Mahalanobis distance thresholds (for accepting or rejecting the claim). *FRR* represents the likelihood of a legitimate request from the true user is rejected. *FAR* represents the likelihood of illegitimate access requests (from imposter) is granted. Two standard metrics were calculated from the DEC: the area under the curve (*AUC*) and the equal error rate (*EER*). *AUC* refers to the area under the *DEC* [7] and *EER* refers to the point on the *DEC* curve where *FAR* is equal to *FRR* [7]. The lower the *AUC* and *EER* value the better the biometric performance.

‘A code’ in the current framework is a gesture performed by a user (either the authorized user or an imposter). Therefore, to facilitate subsequent presentation and discussion, the two words ‘code’ and ‘gesture’ are used equivalently. This nomenclature is also followed for the analysis of the identification system. In the current study, the code length was fixed at one gesture. All the codes were evaluated individually. Assuming a user uses a specific gesture for authentication, which is called the authentication gesture, three possible scenarios would arise in the authentication system: 1) Normal Test, in which the authentication gesture was known only to the authorized user, and the imposter attempts to guess the authentication gesture and gain access, i.e. the code was a secret and not compromised; 2) Leaked Test, in which the authentication gesture was known to an imposter who tries to use it to gain access, i.e. the code was compromised; and 3) Self Test, in which the authorized user forgets the authentication gesture and tries to gain access by performing other known gestures to the system. In all three scenarios, the genuine data is the data from the authentication gesture of the authorized user. For the Normal Test scenario, the imposter data was the data from all the other gestures and all the other users. For the Leaked Test scenario, the imposter data was the data from the authentication gesture of all other users. Lastly, for the Self Test scenario, the imposter data was the data from the other gestures of the true user.

3.2.6 Performance Evaluation of the Identification System

In the literature, the biometric performance of the identification system was quantified by the *rank-k* error [44]. The *rank-k* error represents the likelihood of the authorized user is not among the top k users returned by the identification system. In this study, the smaller the similarity score is, the higher it would be ranked by the identification system. The *rank-k* error is a decreasing function with respect to the value of k as a property of the cumulative match characteristic curve, which plots *rank-k* error against k . In this study, *rank-1* error (R1E), which is a crucial index when only one code is used, was reported for this study. Additionally, the *rank-5* error (R5E) was reported for comparison with previous studies [44]. The cumulative match characteristics (*CMC*) curve was plotted for the optimal configuration of EMG system.

3.2.7 Statistical Analysis

The purpose of this study was to determine the best feature extraction method and the optimal number of channels for both the biometric systems (authentication and the identification) based on EMG. The non-parametric Kruskal Wallis test was employed to determine if the two factors, feature extraction methods (five levels, *i.e.* FDT, TD, and AR, TD+FDT, TD+AR) and the number of channels (eight levels, *i.e.* from 1 to 8), have significant effects on the performance metrics, *EER*, *AUC*, *R1E*, and *R5E*. In case of significance in either of the two factors, the level of the other factor was fixed, and the Wilcoxon rank-sum test was performed on the other factor. For the optimal configuration, performance metrics were reported for the best feature extraction method and the optimal number of channels (for both authentication and identification). The performance of optimal parameters were then compared with the NinaPro DB7. All statistical tests were performed using RStudio 1.0.136 (RStudio, Boston, MA).

3.3 Results

3.3.1 Performance Evaluation of authentication System

Figure 3.3 shows the distribution of the median *EER* and the median *AUC* across all participants, respectively for the three feature extraction methods (FDT, TD, and AR), their combinations (TD+FDT and TD+AR), and a different number of channels (1-8). Median *EER* for the FDT method and for eight channels was 0.76% (Q1=0.1%, Q3=2.93%) for the Normal Test, 2.8% (Q1=0.3%, Q3=5.45%) for the Leaked Test, and 4.86% (Q1=2.1%, Q3=9.89%) for the Self Test (where Q1 is the 25th percentile and Q3 is the 75th percentile). While using a TD feature extraction method for the eight channels, the median *EER* reduced to 0.27% (Q1=0.002%, Q3=2.47%) for the Normal Test and 0.48% (Q1=0.001%, Q3=3.51%) for the Leaked Test. The Self Test using the TD feature set resulted in a median error of 6.68% (Q1=2.76%, Q3=14.04%). The combined feature sets along with TD feature set are illustrated in Figure 3.4. For the Leaked Test scenario, it was observed that the median *EER* for TD+FDT and TD+AR was 0.16% (Q1=10-4%, Q3=0.25%) and 0.18% (Q1=10-4%, Q3=0.25%). The *AUC* followed a very similar pattern to the *EER* for scenarios (reported in the next paragraph). When the number of channels was higher than four, the authentication system's *EER* reached a plateau. These values are similar to other authentication systems reported in the literature [44]. For the three feature extraction methods (FDT, TD, and AR) the initial channel ($n=1$) EER_r was higher than 0.05 for all the scenarios. The EER_r using the TD and FDT method decreased with an increasing number of channels and it reached a plateau ($EER_r < 0.01$) when the number of channels was higher than two. A similar

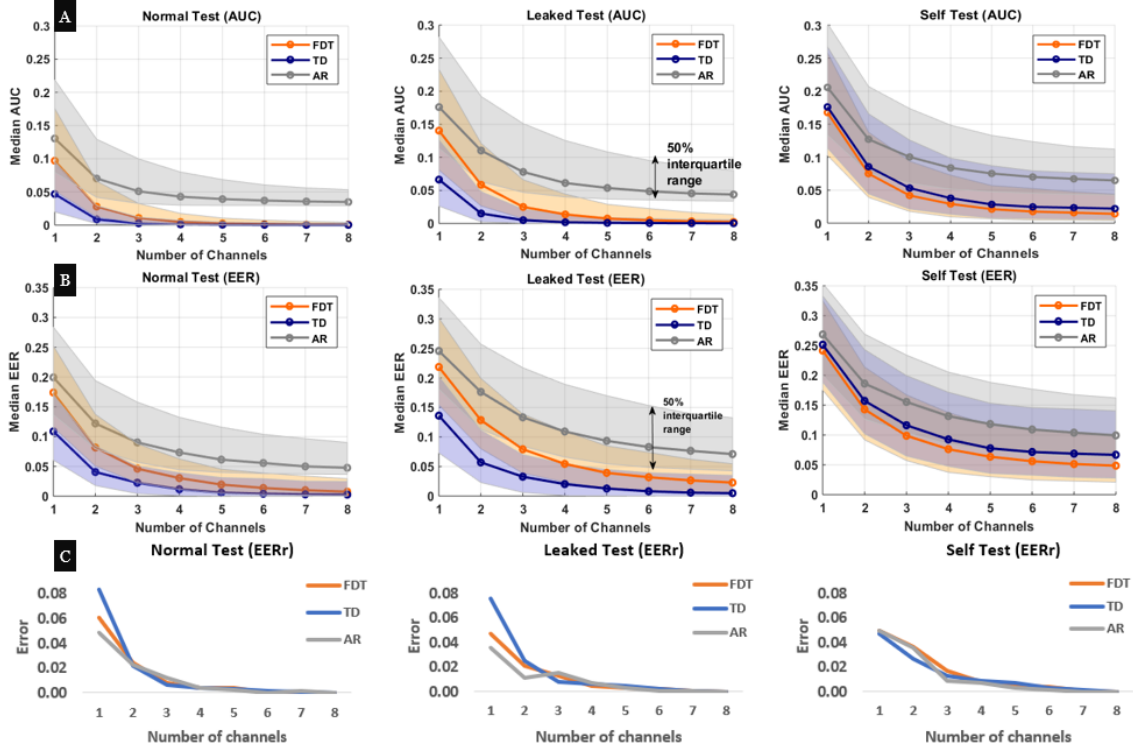


Figure 3.3: (A and B). The median AUC and the median EER for the three feature extraction methods (FDT, TD and AR) and number of channels (1-8) for the Normal Test (left column), Leaked Test (middle column) and Self Test (right column) scenario of the authentication system. Each line indicates the respective median values, and shaded area indicates the corresponding middle 50 percentile (25% to 75%). (C) The range of EER values (EER_r) for the feature extraction methods (FDT, TD and AR) and number of channels (1-8) for the Normal Test, Leaked Test, and Self Test scenario of the authentication system. The range is calculated as a part of the sequential forward selection algorithm at the end of every iteration.

plateau ($EER_r < 0.01$) was reached for the AR method with the number of channels higher than three ($n > 3$). The combined feature sets (TD+FDT and TD+AR) had similar EER_r to the TD feature sets for all number of channels ($n=1-8$).

Feature Extraction Method: In the Normal Test scenario, the median EER was significantly lower for TD than both FDT and AR methods ($p < 0.05$), Figure 3.3 (A)) regardless of channel numbers. The AR method had the highest AUC than the other two methods ($p < 0.05$). Similarly, the median AUC for the TD method was significantly lower than FDT, which was in turn

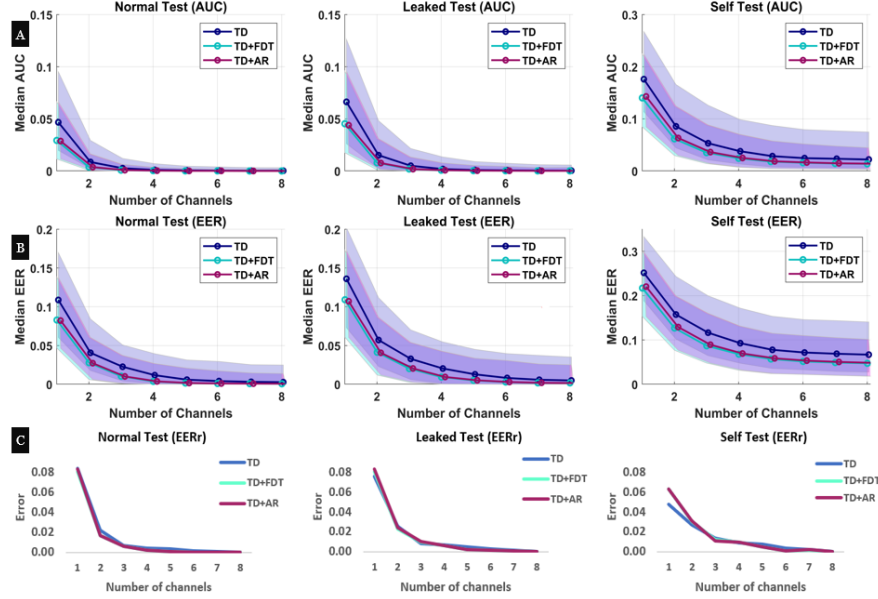


Figure 3.4: (A) The median RIE and the median $R5E$ for the processing methods (FDT, TD, AR) and the number of channels (1-8) for the identification system. Each line indicates the respective median values, and shaded area indicates the corresponding middle 50 percentile (25% to 75%). (C) The range of the RIE (RIE_r) and $R5E$ ($R5E_r$) for the feature extraction methods (FDT, TD and AR) for different number of channels (1-8) for the identification system. The range is calculated as a part of the sequential forward selection algorithm at the end of every iteration.

significantly lower than AR ($p < 0.05$). In the Leaked Test, the EER and AUC both had similar outcomes as the Normal Test: TD outperforms FDT, which in turn outperformed AR ($p < 0.05$, Figure 3.3 (B)). For Self Test, the EER and AUC were not significantly different between the TD and FDT methods ($p > 0.05$, Figure 3.3 (C)) regardless of the channel numbers. However, other than the single-channel case, the EER and AUC of AR were significantly higher than the other two methods ($p < 0.05$). For all three scenarios, the EER_r of the TD method was higher than FDT and AR for lower channel lengths (1-2) (Figure 3.3 (C)). The EER_r was very similar for all the methods for channel lengths (3-8).

Number of Channels: As expected, there was a general decreasing trend in both EER and AUC values with increasing channel numbers (Figure 3.3) in all the scenarios. For the Leaked Test, the decrease in AUC was significant from one to four channels ($p < 0.05$), for all feature extraction methods. The decrease was not significant when more channels were included (5-8, $p > 0.05$). The EER had similar results, with a significant decrease in the lower number of

channels (1-4, $p < 0.05$) for FDT and AR methods. However, for the TD method, the decrease in EER was only significant from one channel to three channels ($p < 0.05$, Figure 3.3 (A)). Similar outcomes were observed in the Normal Test scenario for all three methods. When the channel numbers were smaller, there was a significant decrease with increasing channel number: the AUC and EER decreased significantly for smaller channel numbers (1-4, $p < 0.05$) for TD and FDT methods and from one to three ($p < 0.05$) for the AR method. For the Self Test scenario, the AUC and EER significantly decreased when the channel number increased from one to two ($p < 0.05$) for all three feature extraction methods, and the decrease was not significant for the channel numbers (2 to 8, $p > 0.05$). For all the feature extraction methods the EER_r decreased from channel length 1 to 8, consistently across all analysis scenarios (Figure 3.3 (C)). The decrease in range was minimal (EER_r) when the number of channels was between three and eight.

3.3.2 Performance Evaluation of Identification System

$R1E$ and $R5E$ were used to quantify the biometric performance with different feature extraction methods and channel numbers. Figure 3.5(A) shows the distribution of the $R1E$ and $R5E$ for the different methods and the number of channels. Median $R1E$ for the FDT method and eight channels was 2% (Q1=0%, Q3=12%) and Median $R5E$ was 0% (Q1=0%, Q3=0.2%) for the identification system, which was in agreement with a previous study having a similar experimental setup. The TD feature set had a comparatively lower median $R1E$ of 0.2% (Q1 = 0%, Q3= 4.6%) and a similar $R5E$ of 0% (Q1=0%, Q3=0.1%). The AR feature set had a median $R1E$ of 18.8% (Q1=7.4%, Q3=57.9%). For the combined feature sets, the median $R1E$ for TD+FDT was 0.06% (Q1 = 0%, Q3= 2.9%) and median $R1E$ for TD+AR was 0.2% (Q1 = 0%, Q3= 4.6%). It was evident that the $R1E$ reached a plateau (median $R1E \leq 0.05\%$) when the number of channels used was four and higher. Figure 3.5(b) shows the $R1E_r$ and $R5E_r$ for the feature extraction methods and the number of channels. Both the $R1E_r$ and $R5E_r$ decreased with an increasing number of channels. For the TD feature extraction method, the $R1E_r$ for one channel was 0.17 and the $R1E_r$ for two channels was 0.09. Thereafter, the $R1E_r$ was lower than 0.02 and reached a plateau for more than three channels. The FDT had an $R1E_r$ of 0.076 (n=1), 0.072 (n=2) and 0.05 (n=3) and thereafter it plateaued with $R1E_r < 0.015$. The AR had an $R1E_r$ of 0.057 (n=1), 0.045 (n=2) and 0.028 (n=3) and reached a plateau for more than three channels. The combined feature sets (TD+FDT and TD+AR) had $R1E_r$ similar to the TD feature sets for number of channels higher than two (n>2). The $R5E_r$ showed a similar distribution for the feature extraction methods and the number of channels.

Feature Extraction Methods: The $R1E$ was significantly lower for the TD method ($p < 0.05$) than FDT and AR for all the number of channels (n=1-8). Also, the FDT was significantly lower ($p < 0.05$) than AR for all the channels. For the $R5E$ analysis, the difference between the

feature extraction methods (TD, FDT, and AR) was seen for the lower number of channels (1-3). The TD method had a significantly lower $R5E$ for 1-2 number of channels. However, for the inclusion of more channels (3-8), there was no difference between the $R5E$ for TD and FDT, but the AR had a significantly higher $R5E$ than TD and FDT ($p < 0.05$). The $R1E_r$ was higher for the TD method for the number of channels one and two (1-2) and was the lowest from channel length (3-8, [Figure 3.4 \(C\)](#)). The $R5E_r$ was the highest for the TD method for the first channel number and was the lowest from 2-8. The $R1E_r$ was higher for the FDT method than the AR method for a lower number of channels (1-3) and thereafter the $R1E_r$ dropped below 0.02 for both methods.

Number of Channels: The $R1E$ for the TD method decreased significantly for the number of channels one to four (1-4, $p < 0.05$). There was no significant decrease in further addition. For the FDT and AR method, the $R1E$ significantly decreased for the number of channels one to three (1-3, $p < 0.05$). The $R5E$ significantly decreased for all three methods ($p < 0.05$) for channels 1-2. From 3-8, the error was almost constant with the TD and FDT having a median $R1E$ of zero. For all the methods, the $R1E_r$ decreased from the number of channels 1-8, however, the decrease was minimal from channels 4-8. For all the methods, the $R5E_r$ also decreased from the number of channels 1-8, however, the $R5E_r$ plateaued from channels 3-8.

3.4 Discussion

This paper systematically investigated the effects of feature extraction methods and the channel reduction using EMG to achieve optimal biometric performance in the two application systems: user authentication and user identification. The main advantage of using EMG over other biosignals is the high accuracy of dual-mode security with both knowledge-based and biometrics-based information. The ability to use EMG for recognizing accurate gestures allows the user to choose unique gestures as their own password or code. In the authentication system, the Normal Test scenario fully utilizes this dual-mode security of EMG. The Normal Test scenario represents the security system with both the layers of protection (knowledge-based and biometrics-based). In the Leaked Test scenario, only biometric-based protection is available as the code is compromised. And the Self Test scenario further illustrated the unique property of the EMG-based biometric system, in which the authentication code is forgotten but the legitimate user can be still verified by performing other gestures (codes). With a typical combination of four-channels and FTD feature set, the median EER of the Leaded Test (5.4%) was higher than that of the Normal Test (3%), indicating a 2.4% additional protection provided by the user-defined authentication code ([Figure 3.3 \(A and B\)](#)). The nature of the authentication system makes it challenging to get a low error (specifically $R1E$). The genuine user must be always predicted as rank-1 or else

it is considered an error. The median $R1E$ for a four-channel system was 12.4% (FDT feature set), 3% (TD feature set), and 36.4% (AR feature set). Clearly, the TD feature set had the best biometric performance which is discussed in the next section.

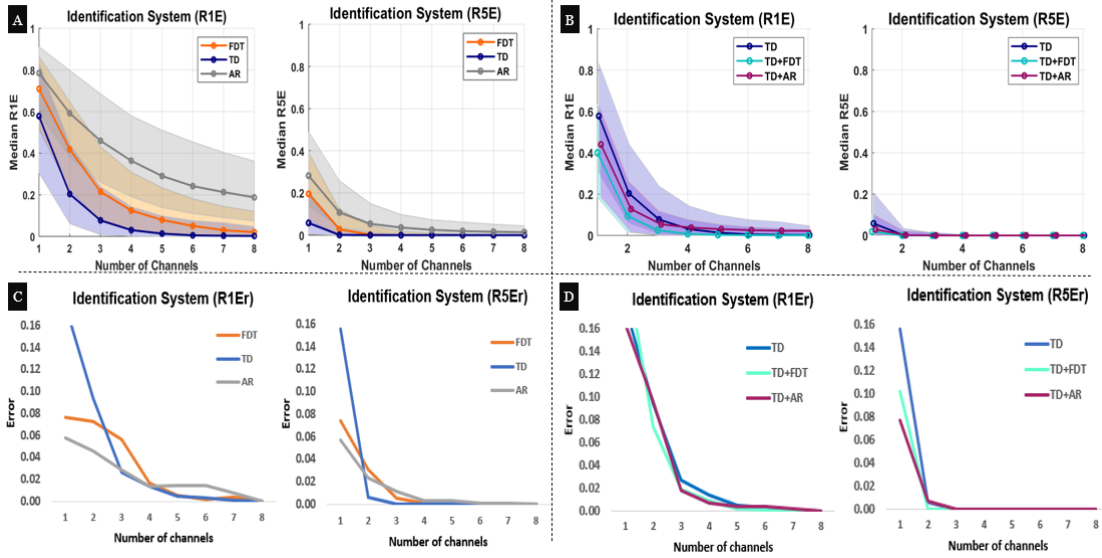


Figure 3.5: (A and B) The median $R1E$ and the median $R5E$ for the processing methods (FDT, TD, AR), their combinations (TD, TD+FDT, TD+AR) and number of channels (1-8) for the identification system. Each line indicates the respective median values, and shaded area indicates the corresponding middle 50 percentile (25% to 75%). (C and D) The range of the $R1E$ ($R1E_r$) and $R5E$ ($R5E_r$) for the feature extraction methods (FDT, TD and AR) and the combined feature sets (TD, TD+FDT, TD+AR) for different number of channels (1-8) for the identification system. The range is calculated as a part of the sequential forward selection algorithm at the end of every iteration

3.4.1 Feature extraction methods

The TD feature extraction method resulted in better biometric performance than FDT and AR for two scenarios of the authentication system (the Leaked Test and Normal Test, shown in Figure 3.3) and the identification system (Figure 3.5). For the Self Test, both the TD and FDT had similar performance across all channel numbers and outperformed the AR method, which had the poorest performance in all number of channels. This was in agreement with the previous studies that compared AR with the TD and FDT methods [99]. For the identification system, the TD

feature set had a significantly lower median *RIE* (3%) than that of the FDT (12.4%). Therefore, the TD feature set has an overall superiority in performance over FDT for both authentication and identification systems. This outcome suggests that the TD feature set is likely more sensitive to individual-differences in EMG, crucial for biometric applications. Previous studies developing a generalized multiuser model found that common spatial-spectral analysis (CSSA) and spectral moments, both similar to FDT, have higher accuracy than the TD features [73,124]. These studies suggest that these features are less user-dependent, which means they capture fewer individual characteristics than TD features, and might explain the finding in this study that the TD feature performing better than FDT and AR feature sets.

While comparing TD with the combined feature sets, the average *EER* of TD (4%) was higher than TD+FDT (2.67%) and TD+AR (2.7%) for the authentication system, and the average *RIE* for TD (3%) was higher than TD+FDT (0.7%) and lower than TD+AR(3.8%) for the identification system. The difference in performance for the TD feature set, although significant ($p < 0.05$), was considered acceptable while comparing to other biometric traits [25]. Due to the superior performance of the TD feature set and the relatively lesser computational complexity than the combined features, it was considered as the optimal feature extraction technique for biometric applications.

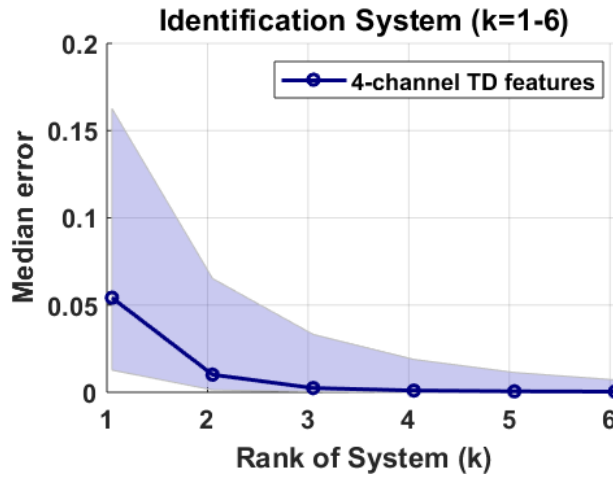


Figure 3.6: *CMC* curve for the identification system using 4 channels EMG and TD Feature extraction.

The median *EER* for the TD feature set and an eight-channel setup was 0.27% (Normal Test), 0.48% (Leaked Test), lower than those reported by a different study (Leaked Test *EER* = 14.96%) which used 64 channels and a combined feature set of (sample entropy, spectral entropy, median frequency, waveform length, and root mean square) [64]. This corroborates our findings

and suggests that TD feature set would have superior performance than other feature extraction methods even with a disadvantage of the number of channels.

3.4.2 Number of channels and channel reduction

As expected, the median EER reduced while increasing the number of channels for all scenarios of the authentication system and for the identification system. This is in accordance with the previous gesture recognition studies that investigated channel reduction [50, 80, 150]. It is evident from Figure 3.3 (A,B) and Figure 3.5 (A,B), a four-channel setup was considered to provide a balance between overall biometric performance and complexity (*i.e.* channel numbers). The error values for the TD feature set are comparable to those reported in previous EMG biometric-related studies [44,50,64]. A previous study used 64 channel HDEMG [64]. Our results indicated that there might not be a need for high-density EMG as four-channel setup provided similar biometric performance. While a one-channel (on the FDS muscle) configuration was used in [50], the half total error rate (HTER) of less than 90% was not sufficient for any practical biometric system.

The error range parameters EER_r , $R1E_r$, and $R5E_r$ were used in the study to determine if the channel location has a significant impact. For all the scenarios it was observed that the first electrode has a higher error range than the remaining channels as shown in Figure 3.3 (C) and Figure 3.5 (C and D). In the sequential forward selection, the lowest error channel was selected in every iteration. Therefore, for the one channel setup (first iteration) the higher error range suggests a significant impact on the first channel location. As observed in all the scenarios, the electrode placed on the Flexor Carpi Ulnaris (FCU) consistently contributed the most towards the biometric performance. This was in agreement with the study that determined the best forearm muscle to control a soft robotic glove [42]. From channels 2-8, the impact was shown to be lower, thus indicating the remaining electrodes could be placed irrespective of specific muscle locations, a desirable feature in practical applications.

3.4.3 Optimal configuration.

The EER and $R1E$ for the four-channel setup and TD feature extraction are listed in Table 3.1. From the above findings, it can be seen that the combination of a four-channel EMG setup, of which one is attached on FCU, and the TD feature set, the average EER for the Normal Test (1.2%), Leaked Test (2%) and Self Test (9.3%) can satisfy the requirement of a authentication system (shown in Table 3.1). The identification system $R1E$ was 3% for this setup. These values are comparable to another study that reported an error of 0.5%, however, they used 64

channels and a combination of different features [65]. Another study used an eight-channel forearm bracelet and a combination of frequency domain and time domain features and obtained a *FAR* of 0.2% and *FRR* of 2.9% [121]. A different study using an eight-channel setup and FDT feature set reported an *RIE* error of 9% [44]. In the present study, a very low *RIE* of 0.2% can be achieved while using an 8-channel configuration and TD feature extraction. Figure 3.6 presents the cumulative match characteristics (*CMC*) curve for such an identification system. It was observed that with an increase in rank (*k*) of the system, the recognition error (*RkE*) reduced, thus improving the performance. The analysis was repeated for the NinaPro DB7 database consisting of 20 participants and performing 17 hand and finger gestures [76]. It was found that for a similar set of parameters (4 channels and TD feature set) the performance was slightly lower than the present study for the Leaked Test (*EER*=6.4%) and Normal Test (*EER*=6.8%, reported in Table 3.1). The lower performance observed with the Self Test (*EER*=23.5%) and Identification System (*RIE*=10.9%) might be due to the partly isotonic nature of the hand gestures, suggesting further investigation of feature sets specific for dynamic contractions. Therefore, the optimal configuration for an EMG-based biometric system consists of four EMG channels and a TD feature set. If there is a higher performance requirement, using 8-channel setup will provide enhanced performance. The findings of the study helped determine the best feature set in terms of performance accuracy and computational complexity as well the optimal number of channels necessary to make accurate biometric systems using EMG.

Table 3.1: TD Feature Extraction and Four-Channel EMG Configuration

Authentication (<i>EER</i>)	Current Study			NinaPro Database		
	Med	Q1	Q3	Med	Q1	Q3
Normal Test	0.012	0.001	0.040	0.064	0.034	0.112
Leaked Test	0.020	0.001	0.055	0.068	0.032	0.119
Self Test	0.093	0.048	0.172	0.235	0.182	0.297

Identification (<i>RIE</i>)	Current Study			NinaPro Database		
	Med	Q1	Q3	Med	Q1	Q3
	0.030	0.001	0.143	0.109	0.045	0.219

TD: Time domain feature set, *EER*: Equal error rate, *RIE*: Rank-1 Error,
Q1: 1th quartile (25%) Q3: 3rd quartile (75%)

The performance reported in Table 3.1 was comparable to the other commonly used biometric traits (fingerprint, iris, and facial recognition) reviewed in a previous study [25]. The EMG-Biometrics has two similar functionalities as ECG and EEG: hidden nature and liveness

detection. The EMG possesses unique knowledge-based security which allows users to customize their access codes as different gestures. This feature, although not available in ECG can be achieved using EEG by using different mental states for codes. As only accuracy $>70\%$ has been achieved for classifying two mental states [148], this is almost impossible for real-world applications. The Leaked Test authentication which is a measure of the biometric-based security achieved a median *EER* of 2%. This was comparable to the reported *EER* of ECG and EEG biometrics which was 0.1% to 5% [100], and 1% to 20% [11, 41]. For the Identification system, the median *RIE* was 3% which was in range with the *RIE* reported for ECG and EEG, 0%-20% [100] and 1% to 20% [11, 41], respectively. This suggests that the four-channel EMG with TD feature extraction can achieve comparable if not better biometric performance than other biosignals.

3.5 Chapter Conclusions

The TD features had improved and better biometric performance than FDT features and AR features for the Leaked Test and Normal Test of authentication system and for all number of channels ($n=1:8$). The TD features had better performance than the FDT feature set in the identification system and for all number of channels ($n=1:8$). For all the feature sets, four channels had a plateau in performance for a higher number of channels for both the authentication and identification systems. Therefore, the TD feature set and the four-channel EMG configuration is found to be ideal for optimal biometric performance using EMG. The range of error was high for the first channel and reduced for the inclusion of further channels. This indicates that one location on the FCU is important and the remaining could be placed irrespective of muscle location. This finding motivates the design of an easy-to-wear arm-band with four equally spaced electrodes and the placement is recommended in a way that one of the electrodes is right on top of the FCU for best biometric performance. Therefore, an accurate EMG-based biometric system can be designed using a TD feature extraction method and four equi-spaced electrodes with one of them positioned directly on the FCU muscle.

Chapter 4

Score, Rank, and Decision-Level Fusion Strategies of Multicode Electromyogram-Based Authentication and Identification Biometrics

This chapter was published in the *IEEE Journal of Biomedical and Health Informatics*, vol 26 as Pradhan, A., He, J., & Jiang, N., “Score, Rank, and Decision-Level Fusion Strategies of Multicode Electromyogram-Based Verification and Identification Biometrics”, 2021 [106]. This chapter presents the development of a multi-code framework for combining hand gestures as passwords. As a note to the reader: the term ‘verification’ in the published document has been interchanged with ‘authentication’ to align with the remaining parts of the thesis.

4.1 Chapter Introduction

Biometrics are biological information or characteristics that can be used to identify an individual or verify an individual’s identity. The two most-used applications of biometric systems are: 1) authentication and 2) identification [7]. Authentication is when the system uses biometric information to accept or reject a claim of being a particular user by comparing the biometric information presented by the claimant to the target user’s template stored in the database. Conversely, identification is when the system tries to infer the identity of a claimant by finding the closest match in the database. Recently, biosignals such as electroencephalography (EEG)

and electrocardiography (ECG), which represent the electrical activity of the brain and heart, have been proposed as biometric traits [11, 41, 100]. The endogenous and latent nature of these biosignals makes it difficult to acquire and synthesize compared to other commonly used biometrics such as fingerprints, iris, and facial features. Electromyography (EMG) is another type of electrophysiological signal that has been extensively used for accurate hand gesture recognition [28, 46]. Currently there exists difficulties in realizing such a one-size-fits-all EMG gesture recognition model [73, 89, 124]. This is indicative of the presence of individual differences in the EMG signals, and hence, possess great potential as a biometric trait.

Indeed, several recent studies have analyzed the performance of upper extremity EMG biometrics as a biometric trait [44, 62, 64, 65, 82, 121]. A unique feature of EMG in the context of biometrics is that it allows an inherently dual-mode security: 1) knowledge (KW) security, where the user can set customized hand gestures as codes, and 2) biometrics (BM) security, the nature of the EMG that can identify individuals. The KW+BM security allows the user to customize their code sequence as a combination of multiple codes as desired. The concept of using a gesture sequence as a neuromuscular password was introduced in a previous study [64]. A different authentication study allowed users to choose three out of four gestures as their authentication code [121]. An identification study with a similar setup as in [64] analyzed the biometric performance across multiple codelengths [62]. It was shown that with increasing codelength, the biometric performance increased. However, the framework of a multicode EMG biometric system, in which the maximum potential of dual security is realized, has not been systematically analyzed for authentication or identification applications.

Therefore, the aim of the study was two-fold: 1) to compare the biometric performances of the three fusion strategies and 2) to determine the desirable parameter (codelength and rank-K) values for the three fusion methods. To the best of our knowledge, this study is the first of its kind to investigate the development of a multicode EMG biometric system and its performance in multiple authentication and identification scenarios.

4.2 Methods

4.2.1 Participants

We recruited 24 healthy participants (13 males and 11 females) for the study. The age of all the participants was between 19 years and 30 years and the average forearm length (measured from the styloid process on the wrist to the olecranon on the elbow) was $25.4 \text{ cm} \pm 1.46 \text{ cm}$. The participants signed an informed consent form, which allowed them to withdraw at any point

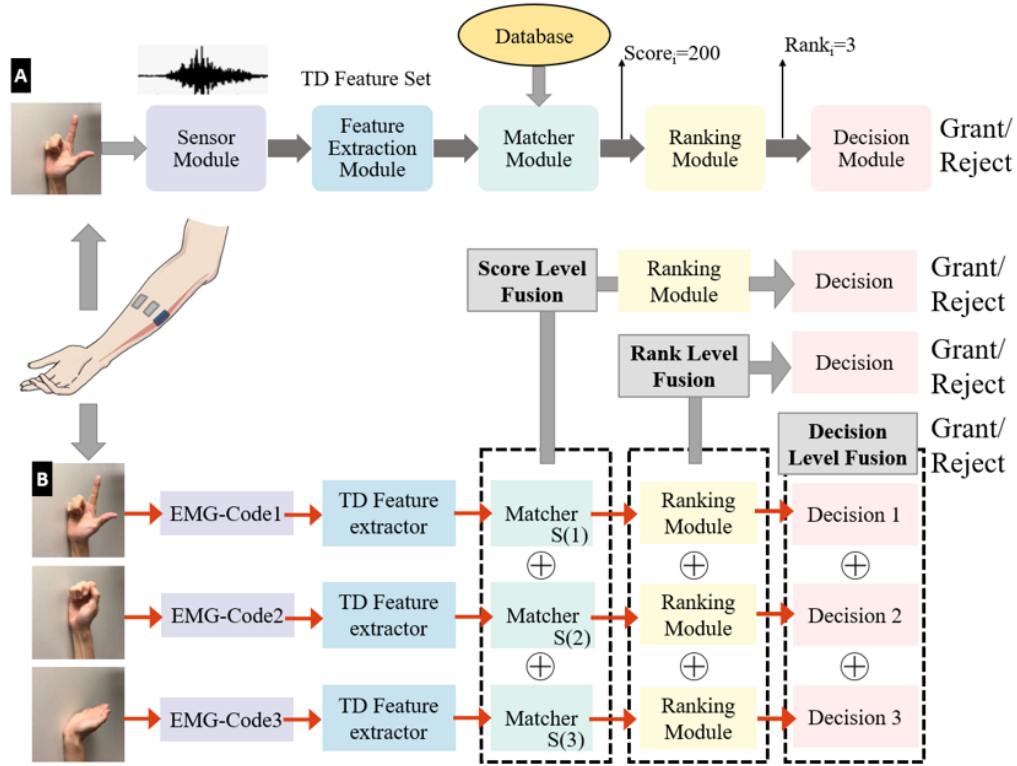


Figure 4.1: A) Unicode EMG Biometric system. The processed input is matched with the template(s) in the database and a score is generated for every matching operation. B) Multicode EMG biometric system combines individual codes using a post-matching fusion at one of the three levels—score, rank, and decision. For both the systems, the ranking module is present only for identification, whereas the authentication application bypasses the ranking module, and a decision is made based on the score threshold.

in the study. The experimental protocol was approved by the Office of Research Ethics of the University of Waterloo (ORE# 22391). Additionally, data from 60 participants (age: 22-45), publicly available through the combination of NinaPro DataBase DB2 [4] and DB7 [76], was used for the comparison of results.

4.2.2 Experimental Protocol

Eight monopolar EMG electrodes (AM-N00S/E, Ambu, Denmark) were placed in the form of two rings, each consisting of four electrodes equally spaced around the forearm and forming four bipolar pairs. Furthermore, the differential of each pair was used for subsequent processing, as shown in Figure 4.2. To maintain uniformity across all participants, one electrode in each ring was anatomically positioned on the flexor carpi ulnaris muscle [105]. The distance between the two rings was 2 cm and the proximal ring was at a distance of one-third the forearm length measured from the elbow crease. The participants were seated comfortably on a height-adjustable chair, and both their upper limbs were in a resting position. A computer screen was placed in front of the participants, which displayed visual instructions for performing the gestures. A commercial EMG amplifier (EMG USB2+, OT Bioelettronica, Italy) was used to acquire the EMG signals (the sampling rate was 2048 Hz, with bandpass cut-offs at 10 Hz and 500 Hz). The following 16 hand and wrist gestures were included in the current study (presented in Figure 3.2): Lateral prehension (LP), thumb adduction (TA), thumb and little finger opposition (TLFO), thumb and index finger opposition (TIFO), thumb and little finger extension (TLFE), thumb and index finger extension (TIFE), index and middle finger extension (IMFE), little finger extension (LFE), index finger extension (IFE), thumb extension (TE), wrist flexion (WF), wrist extension (WE), forearm supination (FS), forearm pronation (FP), hand open (HO), and hand close (HC). Each gesture was repeated three times and five seconds of recording was performed in each repetition. To avoid EMG signals from transitory movements, the participants were asked to start the contraction slightly earlier than the beginning of the recording. A five-second rest period was provided between two consecutive contractions to avoid muscle fatigue. The order of the gestures was randomized for each participant. For the public dataset (NinaPro DB2 and DB7), participants performed 17 isometric and isotonic, hand and finger gestures. Six repetitions of each gesture were performed and 5 seconds of data were recorded using a Delsys Trigno IM Wireless EMG system [4, 76].

4.2.3 EMG Signal Processing

The bipolar EMG signals were first windowed with a window length of 200 ms and an overlap of 150 ms [105]. Each window was then processed using the Hudgin’s time domain (TD) feature extraction technique [52]. The TD feature set was employed based on the finding of our recent study [105]. For each epoch of N data-frames, the signal amplitude x_i , the TD feature vector included the root mean square (RMS), zero crossing (ZC), slope sign changes (SSC), and waveform length (WL) as shown:

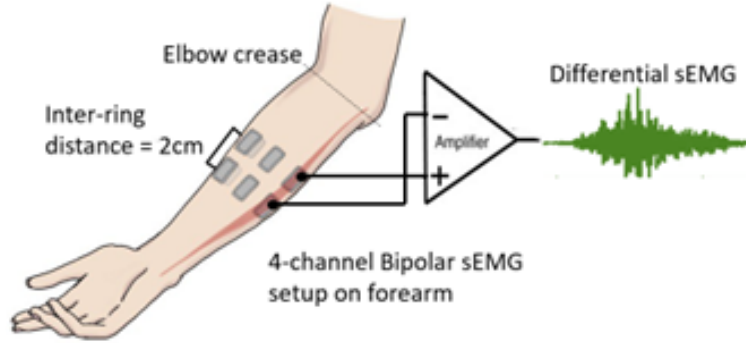


Figure 4.2: Positions of the eight surface electrodes on the forearm (dorsal view, two electrodes are not visible as they are in the posterior side). These eight electrodes formed four bipolar pairs in the proximal-distal orientation and the differential EMG was used for subsequent signal processing. The first pair is placed on the flexor carpi ulnaris (FCU) muscle and the remaining three are placed uniformly around the forearm.

$$\text{RMS} = \sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2} \quad (4.1)$$

$$\text{SSC} = \sum_{i=2}^{N-1} f(x_i - x_{i-1})(x_i - x_{i+1}) \quad (4.2)$$

$$f(x) = \begin{cases} 1, & \text{if } x \geq T_h, \\ 0, & \text{otherwise.} \end{cases}$$

$$\text{WL} = \sum_{i=1}^{N-1} |x_{i-1} - x_i| \quad (4.3)$$

$$\text{ZC} = \sum_{i=1}^N \text{sgn}(x_i \times x_{i+1}) \cap |x_i - x_{i+1}| \geq T_h \quad (4.4)$$

$$\text{sgn}(x) = \begin{cases} 1, & \text{if } x \geq T_h, \\ 0, & \text{otherwise.} \end{cases}$$

where T_h is the threshold value defined by the baseline or resting stage. For a given feature vector sample p , which is the input from a specific user (the claimant) while performing a specific gesture, its matching score, $S_{i,j}$, with the i th gesture and the j th user, was defined as the Mahalanobis distance between the sample and the class centroid:

$$S_{i,j}(p) = \sqrt{(p - \mu_{i,j}) \Sigma_{i,j}^{-1} (p - \mu_{i,j})^T} \quad (4.5)$$

Where $\mu_{i,j}$ is the centroid of the gesture of the class and the user, and $\Sigma_{i,j}$ is the covariance matrix for the specific gesture and user class. Both the centroid and covariance were estimated from the enrollment data. In the current study, two trials of each gesture were used as enrollment data (training) and the remaining trial was used as claimant data (testing), resulting in a leave-one-out (LOO) cross-validation scheme. The biometric performance of each fold was quantified using the metrics as described in the upcoming part.

4.2.4 Multi-code Performance Evaluation

The *FAR* and *FRR* for the multi-code configuration were calculated separately for the two biometric applications, authentication and identification. The performed gestures (also termed as codes to facilitate the subsequent presentation) generated the biometric information at each code. An example of a user's authentication code sequence of codelength M would be $C_1, C_2, \dots, C_m, C_M$, in which C_m is the m^{th} code or gesture. For this study, hundred random sequences ($n=100$) of codelength six ($M = 6$) were generated. To assess the effect of codelength for these sequences, the *FRR* and *FAR* metrics were used to compare the fusion techniques by varying the codelength from one to six ($M = 1, 2, 3, \dots, 6$) for each code sequence. For the identification mode, an additional parameter, rank- K , was analyzed by varying the value of K ($K = 1, 2, 3, 4$) at different codelengths for each code sequence.

Performance Evaluation of Authentication mode

For the biometric authentication mode, the detection error tradeoff curve (*DET*) is the relationship between the *FAR* and *FRR*. The equal error rate (*EER*) is the point on the *DET* curve where the *FAR* is equal to the *FRR*. The *EER* is a commonly used authentication metric that can be used to compare the performance of different biometric traits: The lower the *EER* value, the better the performance. To investigate the performance of the KW and the BM aspects of a multicode EMG biometric system, two authentication scenarios were studied: 1) a normal test where the authentication gesture was only known to the genuine user, while the impostor had no knowledge

of the authentication gesture and must guess; 2) the leaked test where the authentication gesture specific to the genuine user is known to the impostor who tries to use it to gain access. In both cases, for the m^{th} code, the genuine score, G_m , was obtained from the authentication gesture C_m of the genuine user. In the normal test, the impostor score I_m was obtained from the other gestures performed by other users. The leaked-test scenario I_m was obtained from the gesture C_m of all the other users. Fusion schemes were implemented using G_m and I_m to obtain the final FAR_M , FRR_M , and EER_M .

4.2.5 Performance Evaluation of Identification mode

For a rank- K biometric identification, the recognition accuracy of the system $Accuracy_{m,k}$ is defined as the probability of correctly predicting the genuine user within the top K likely users returned by the system at the m^{th} code. The likelihood of the genuine user not being ranked in the top K is defined as

$$FRR_{m,k} = 1 - Accuracy_{m,k} \quad (4.6)$$

It was expected that increasing the value of K would decrease $FRR_{m,k}$ but consequently allow more impostors in. Therefore, $FAR_{m,k}$ in a rank- K identification was defined as the probability of impostors being ranked within the top K likely users at the m^{th} code. Multiple fusion schemes were used to develop the multicode system and were compared with their respective $FRR_{m,k}$, and $FAR_{m,k}$. Furthermore, the relationship between the parameters K and M was analyzed, and the desirable parameters for an EMG biometric system were established.

4.2.6 Fusion Schemes

In this study, fusion schemes at three levels were compared for a multi-code biometric system: score-, rank-, and decision level. Note that the rank-level fusion does not apply to the authentication mode, while all the fusion schemes are applicable to the identification mode. For the following sections, a code sequence of codelength M is denoted as $C_1, C_2, C_m, \dots, C_N$ in which C_m is the m^{th} code or gesture.

Score-level fusion

One of the most common score-level fusion [119] is the simple summation of the scores of multiple matcher modules. It was expected that these scores from multiple codes would vary in their respective distributions, therefore, normalization was performed prior to the summation step

to address these variations [68]. Several normalization methods have been discussed in previous research [67]. Owing to differences in the analysis, two separate normalization approaches were considered for the authentication and identification. For the authentication mode, the mean-to-overlap extrema-based anchored min-max normalization was chosen, because it takes into consideration the overlap region between the genuine scores G_m and the impostor scores I_m for the m^{th} code (shown in Figure 4.3 (A)). The anchor value of the overlap region was calculated as follows:

$$A(r, n) = \{\max(G(m)) - \mu(G(m))\} + \{\min(I(m)) - \mu(I(m))\} \quad (4.7)$$

where $\mu(G(m))$ and $\mu(I(m))$ are the mean of the genuine and impostor scores, respectively. The anchor value was further utilized to obtain the normalized score:

$$s_n(r, n) = \begin{cases} \frac{s(m) - \min\{G(m), I(m)\}}{2A(m) - \min\{G(m), I(m)\}}, & \text{if } s(m) \leq A(m), \\ 0.5 + \frac{s(m) - A(m)}{\max\{G(m), I(m)\} - A(m)}, & \text{if } s(m) > A(m). \end{cases} \quad (4.8)$$

For identification, a w-score normalization was employed [122]. Instead of following the central limit theorem where the median values are used to analyze a distribution, an extreme value theory (EVT) is adopted which focuses on the extreme values as these values will most likely include the scores of the genuine user. The w-score normalization is an extreme value analysis where the top impostor scores are used to fit a Weibull distribution (Figure 4.3 (B)). This is suitable for identification purposes, as the single best representation of each template score set was used for comparison. The Weibull distribution could be fitted with as low as six data points [122]. The normalized scores were generated using the cumulative distribution function (CDF) of the Weibull distribution:

$$s_n = CDF(s) = 1 - e^{-\left(\frac{s}{\lambda}\right)^k} \quad (4.9)$$

where k and λ are the shape and scale parameters estimated from the Weibull distribution, respectively. For both the authentication and identification modes, the extrema-based normalization was termed as ‘Score-Norm.’ For comparison purposes, the simple summation with min-to-max normalization was also analyzed (Score-MM).

Rank-level fusion

In a rank- K identification, the output of each code is a vector, i.e., $r = [r_1, r_2 \dots r_j \dots, r_J]$, where r_j is the rank of the j^{th} enrolled user to be the user currently presenting biometric information. In a single-code case, access is granted to K users with the highest probabilities. In the

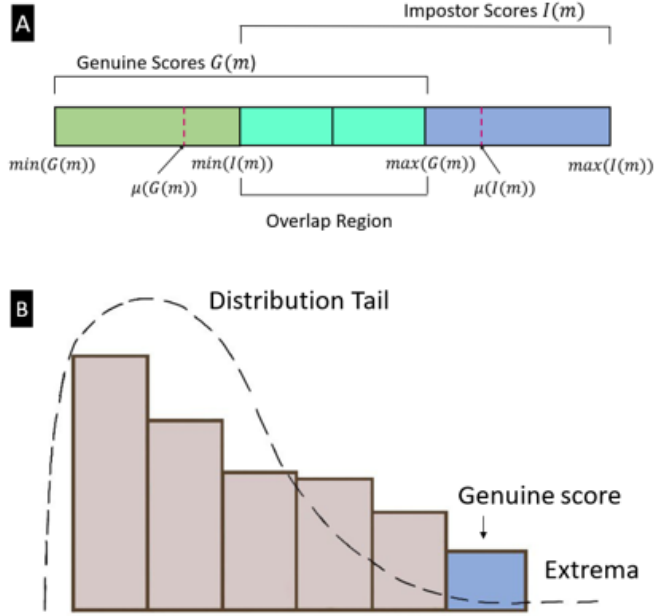


Figure 4.3: A) The genuine and impostor score sets as the output of a matcher. The anchor value is calculated based on the overlap region and the mean values of the genuine and impostor score set [105]. B) The identification involves sample by sample comparison of the presented user to the template database. The outcome contains a genuine score (blue), the rest are impostor scores (brown). The best case is when the genuine score shows up in the tail of the score distribution. The best score in the tail is (in case of a match) an outlier with respect to the Weibull distribution model of the impostor scores [52].

multi-code case, the rank-level fusion combines the rank outputs of all codes using two methods, the Borda count and logistic regression methods [91], to provide an enhanced level of precision. The Borda count method (Rank-Borda) uses the summation of the ranks generated by individual codes to calculate the final rank. For M codes, the statistic (st) computed for each user ‘ j ’ was:

$$st_j = \sum_{m=1}^M r_{m,j} \quad (4.10)$$

The statistic for each user determines the final rank of the identification mode such that the user with the lowest value of st is assigned the highest rank. The logistic regression method (Rank-LR) uses a weighted sum of the individual codes to calculate the final rank [120]. The weight to be assigned to different codes (w_m) was calculated using logistic regression where two

trials of each gesture was used for training and one trial for testing and a 3-fold cross-validation was performed. The individual gesture performances are listed in [Appendix A](#). For M codes, the statistic computed for each user was:

$$st_j = \sum_{m=1}^M w_m r_{m,j} \quad (4.11)$$

where w_m is the weight of the m th code.

Decision-level fusion

In a multi-code biometric system, the simplest decision-level fusion method is obtained by using the logical AND rule on the outcomes of individual codes (Decision-AND). The output of the Decision-AND is a correct match when all the codes agree that the user presenting biometrics matches the template. Another method is to use majority voting. However, majority voting does not account for the individual performance of a code, and it can result in a stalemate with an even codelength. To address these limitations, weighted majority voting (Decision-WM) [120] was used in the current study. For the authentication mode, the presented user was compared to the corresponding template in the database. Hence, it was a binary problem with outputs of 1 (for the genuine user) and 0 (for an impostor). For identification, it was a multiclass problem with J comparisons between the presented user and all the templates in the database, however, only the top K users were granted access. For M codes, let $d_{m,j}$ be the degree of support for the j^{th} user, defined as:

$$A(r, n) = \{\max(G(m)) - \mu(G(m))\} + \{\min(I(m)) - \mu(I(m))\} \quad (4.12)$$

The discriminant function for each user j obtained through the weighted voting is:

$$g_j = \sum_{m=1}^M w_m d_{m,j} \quad (4.13)$$

where w_m is the weight attached to the m^{th} code. Here, w_m is the single-code accuracy of each code normalized to one:

$$\sum_{m=1}^M w_m = 1 \quad (4.14)$$

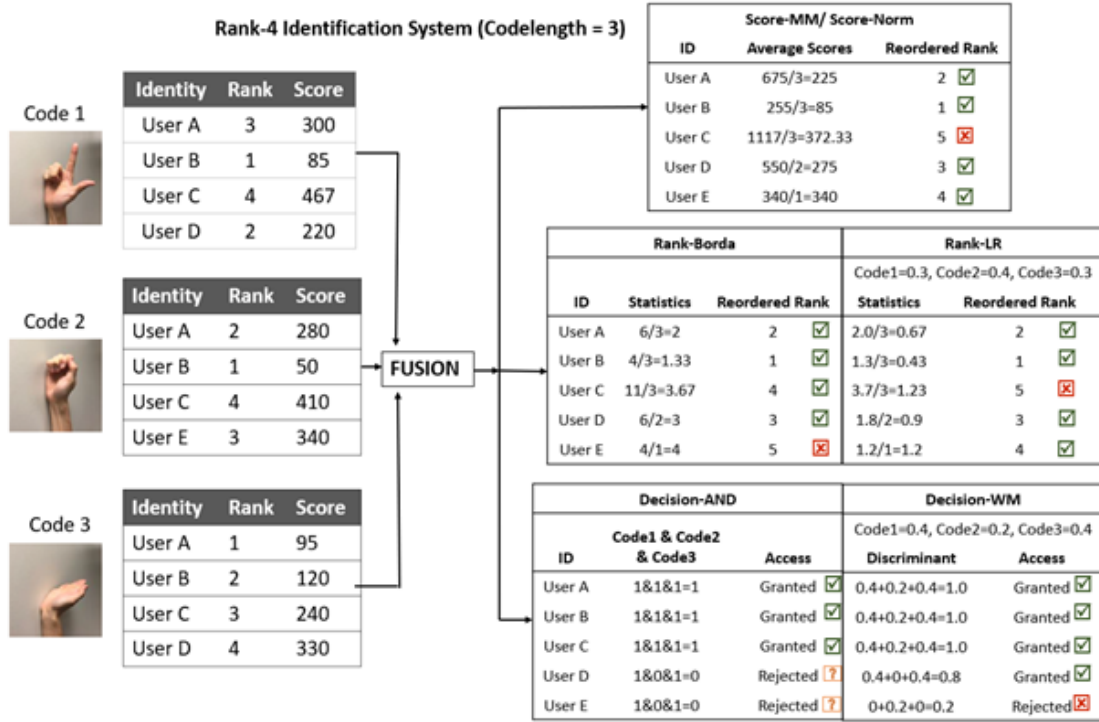


Figure 4.4: The working of a multimodal EMG identification mode. A scenario with five users and three codes are presented, and different fusion schemes are performed. In a rank-4 system, four users are granted access. In the case of less than four users qualifying the selection criteria, the remaining users are chosen randomly (as seen in Decision-AND, one of user D or user E is selected at random)

For the authentication mode, the presented user matches the template if the discriminant g has a majority ($> 50\%$). For the identification mode, the presented user was assigned to the template with the highest value of g . The weight w_m was the previously determined performance accuracy of the m^{th} code (gesture) normalized to one. The individual gesture performances are listed in [Appendix A](#). The working of the different fusion schemes for the identification mode is further illustrated in [Figure 4.4](#).

4.2.7 Statistical Analysis

The aim of the study was to compare multiple fusion schemes and to determine the desirable parameter values in a multi-code EMG biometric system in two applications, authentication and

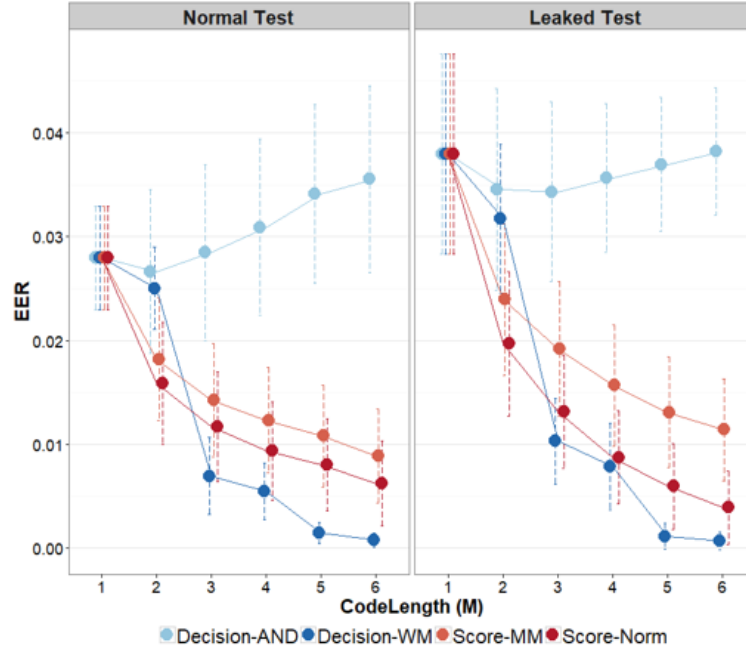


Figure 4.5: Mean EER for the multicode authentication mode scenarios (normal test and leaked test). The codelength (M) was varied from 1 to 6 and different fusion schemes were employed: Decision-AND, Decision-WM, Score-MM, and Score-Norm. The dashed error bars represent the standard deviation.

identification. An ANOVA was used to determine if there was any significant effect of fusion schemes (six levels, i.e., Score-MM, Score-Norm, Rank-Borda, Rank-LR, Decision-AND, and Decision-WM), and codelength (six levels, i.e., from 1 to 6) on the EER of the authentication mode. For the identification mode, an additional factor of rank- K (with four levels, i.e., from 1 to 4) was added to the ANOVA model to test their effect on the FAR and FRR metrics. In the case of significance in any of the factors, the levels of the other factor (factors, for identification) were sequentially fixed and the Bonferroni post-hoc analysis was performed on the significant factor. The recommended parameter values were determined from the lowest error rates. All statistical tests were performed using RStudio 1.0. 136 (RStudio, Boston, MA).

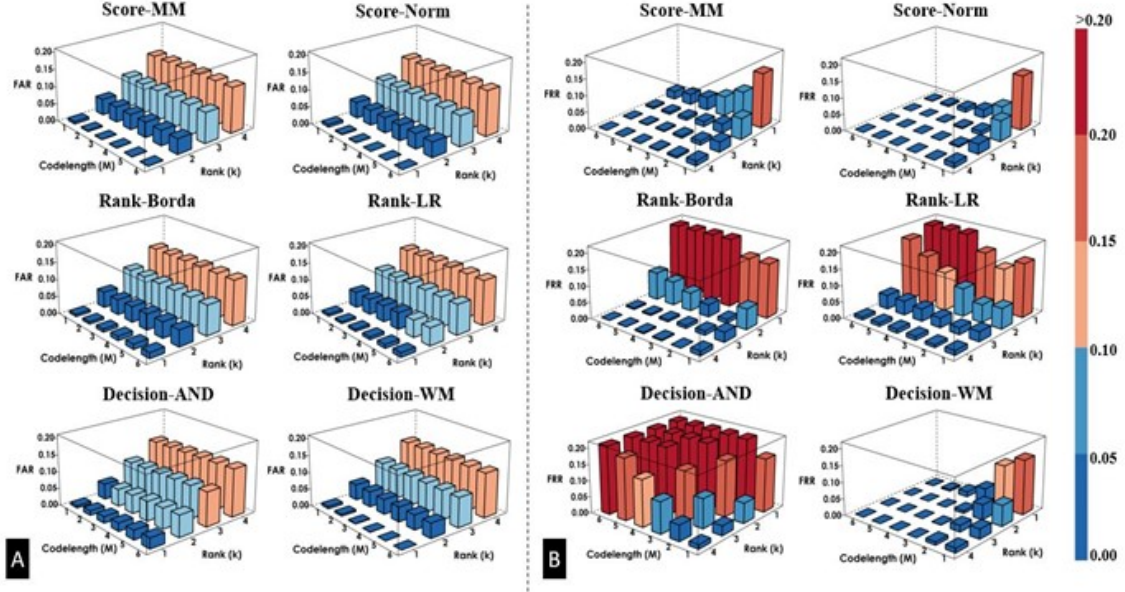


Figure 4.6: A) Mean *FAR* for varying codelength (*M*) and rank (*K*) of the multimodal identification mode using different fusion schemes. B) Mean *FRR* for varying codelength (*M*) and rank (*K*) of multimodal identification mode using different fusion schemes. All bars are color-coded according to the legend on the right. Note that the purpose of presentation clarity, the bars with values higher than 0.2 have been clipped at 0.2.

4.3 Results

4.3.1 Performance Evaluation of Authentication Mode

Figure 4.5 shows the mean *EER* values of the different fusion schemes at different codelengths and for the two authentication scenarios, the normal- and leaked test.

1) Fusion Scheme Analysis

In a unicode system ($M=1$), all the fusion schemes had an identical mean *EERs* for the normal test (0.027 ± 0.004) and leaked test (0.037 ± 0.0009), which was expected and in agreement with a similar study previously published [44]. For the normal-test scenario with $M=2$, the *EER* of Score-Norm (0.015 ± 0.005) was significantly lower ($p < 0.05$) than those of Score-MM (0.018 ± 0.005), which in turn was significantly lower ($p < 0.001$) than those of Decision-AND ($0.026 \pm$

0.007) and Decision-WM (0.026 ± 0.004). However, from $M=3$ onwards, the *EER* of Decision-WM became consistently and significantly lower than the other three methods, with 0.007 ± 0.003 , 0.005 ± 0.002 , 0.001 ± 0.001 , and $10^{-4} \pm 10^{-4}$ for $M=3-6$, respectively. Another striking trend was that the *EER* decreased with an increasing number of codes, other than Decision-AND. Similar trends were found for the leaked-test scenario: Decision-WM outperformed the other three methods for $M>2$; Decision-AND's *EER* increased with increasing M , while the *EER* of the other three methods decreased. The lowest *EER* values of Decision-WM were 0.010, 0.004, 0.007 ± 0.004 , 0.001 ± 0.001 , 10^{-4} , 0.0008, for $M=3, 4, 5$ and 6, respectively. For the Decision-AND and Score-MM fusion schemes, the normal-test *EER* values were lower than those of the corresponding leaked test under all conditions. For the Score-Norm with $M \leq 3$, the mean *EER* of the normal test was lower than that of the leaked test, while for $M>3$ the opposite results were observed. The Decision-WM's *EER* values for the normal test were lower than those for the leaked test, although for $M=5$ and 6, *EER* values for the two scenarios were very similar (Figure 4.5).

2) Codelength (M) Analysis

From Figure 4.5, it was also observed that the mean *EER* of Decision-WM, Score-MM, and Score-Norm decreased while increasing the codelength for the authentication scenarios. However, for the Decision-AND, the mean *EER* decreased until $M=2$ and thereafter increased until $M=6$. For both authentication scenarios, it was also observed that the mean *EER* of Decision-WM significantly decreased ($p<0.001$) for an increase in the codelength from $M=2-3$ and from $M=4-5$. However, there was no significant difference ($p>0.05$) when the codelength increased from 1-2, 3-4, and 5-6. The mean *EERs* of Score-Norm and Score-MM decreased from $M=1-6$.

4.3.2 Performance Evaluation of the Identification Mode

The performance of identification mode was measured using its *FRR* values (as in (6)). Additionally, the *FAR* values were also calculated to evaluate the identification performance. Figure 4.6 shows the distribution of ($FAR_{m,k}$) and ($FRR_{m,k}$), for the six fusion schemes, as a function of the codelength (M) and rank (K).

1) Fusion Scheme Analysis

As seen from Figure 4.6(B), for all the fusion schemes and codelengths, the *FRR* decreased with an increase in K , which was expected. The unicode *FRR* values for the rank-1 system (the right-

most bars in the plots in Figure 4.6(B)) were equal for all the above fusion schemes (0.165 ± 0.05) as expected. In general, we observed that for all values of K , the mean ($FRR_{m,k}$) of Rank-LR, and Rank-Borda decreased from $M=1$ to $M=2$ and increased again until $M=6$. For all K , the mean ($FRR_{m,k}$) of Decision-AND significantly increased ($p < 0.05$) with an increase in M . As such, Rank-Borda, Rank-LR, and Decision-AND are not suitable for multicode identification. However, it was observed that the $FRR_{m,k}$ of Score-MM, Score-Norm, and Decision-WM decreased with increasing codelength, which was reported as follow: For the rank-1 system ($K=1$), it was observed that ($FRR_{2,1}$) of Score-Norm (0.053 ± 0.032) was significantly lower ($p < 0.001$) than that of Score-MM (0.092 ± 0.047) and Decision-WM (0.131 ± 0.048). Similarly, the mean ($FRR_{3,1}$), of core-norm (0.025 ± 0.017) was significantly lower ($p < 0.001$) than that of Decision-WM (0.043 ± 0.019) and Score-MM (0.060 ± 0.036). However, for $M=4-6$, the mean $FRR_{m,1}$ was not significantly different ($p=0.33$) between Score-Norm ($FRR < 0.012$) and Decision-WM ($FRR < 0.015$), both of which were significantly lower than all the other fusion schemes ($p < 0.001$). For the rank-2 system, the unicode FRR values ($FRR_{1,2}$) were equal for all the fusion schemes (0.062 ± 0.003). The mean ($FRR_{2,2}$) for Score-Norm (0.011 ± 0.011) and Rank-Borda (0.011 ± 0.009) was significantly lower ($p < 0.001$) than the other fusion schemes. However, the ($FRR_{3,2}$) for Score-Norm (0.003 ± 0.006) was significantly lower ($p < 0.001$) than that for Decision-WM (0.008 ± 0.007), which in turn was significantly lower than all the other fusion schemes ($p < 0.001$). For $M=4-6$, the ($FRR_{m,1}$) was not significantly different ($p > 0.05$) between Score-Norm ($FRR < 0.003$) and Decision-WM ($FRR < 0.008$), both of which were significantly lower ($p < 0.001$) than all the other fusion schemes. A similar trend was observed for the rank-3 and rank-4 systems.

2) Codelength (M) and Rank (K) Analysis

From Figure 4.6(B), it was also observed that the mean FRR of the Score-Norm, Score-MM, and Decision-WM methods attained a plateau with very small $FRR_{m,k}$ (~ 0.00). With a rank-2 system, both the Score-Norm and Decision-WM attained a plateau with the mean $FRR_{m,k} = 10^{-4}$ for $M=4-6$. With a rank-3 system, both the fusion schemes attained a plateau with $FRR_{m,k} = 10^{-4}$ for $M=3-6$, while for $K=4$, a plateau with $FRR_{m,k} = 10^{-4}$ was attained for $M=2-6$.

From Figure 4.6(A), for a particular value of K , there was generally no significant difference ($p > 0.05$) between the different codelengths for all the fusion schemes. The only exception was the Decision-AND where the rank-1 $FAR_{m,1}$ increased from 0.007 ± 0.002 ($M=1$) to 0.030 ± 0.001 ($M=6$). It was also found that in general, the $FAR_{m,k}$ of all the fusion schemes increased linearly with respect to K , however, there was no significant difference ($p > 0.05$) between the $FAR_{m,k}$ values for the different fusion schemes.

4.4 Discussion

This study presented the development of a fusion-based multicode EMG biometric system. The dual security (KW and BM) of such a system was evaluated by the normal- and leaked test of the authentication mode. The normal-test facilitates both these securities, whereas in the leaked-test scenario, the code was compromised; hence, only BM security was available. The BM security was further explored by the identification performance, where the *FRR* typically increased by *N*-folds compared to the corresponding authentication metrics (where *N* is the number of users in the database). The results of this study demonstrated the authentication and identification performance of the fusion schemes while varying the parameters of the system, i.e., the codelength (*M*) and rank (*K*), as discussed in the next section.

4.4.1 Performance of Fusion Schemes

In a multicode system, post-matching fusion is possible at three different levels: score, rank, and decision. Therefore, this study analyzed the performance of six different fusion schemes (Score-MM, Score-Norm, Rank-Borda, Rank-LR, Decision-AND, and Decision-WM). The authentication evaluation results showed that for the normal- and leaked test, the *EER* of Score-Norm was significantly lower than Score-MM for multiple codelengths ($M=2-6$). This suggests that the proposed overlap extrema-based normalization is more effective than the traditional min-to-max normalization for a multicode authentication mode. This finding was in accordance with a previous study that compared these normalization techniques for a different multimodal (fingerprint, face, and iris) system [67]. Another interesting observation for the Score-Norm fusion was that the leaked-test *EER* was lower than the normal-test *EER*, which was contradictory to previous findings [44], where the EMG-Biometrics provided a higher *EER* for the leaked-test. Although this compromises the KW security of the system, the BM security was highly improved. This is advantageous in real-world scenarios, where the code was stolen by an impostor. The Decision-AND had a relatively poorer performance because the *EER* was observed to increase for $M>3$ for both authentication scenarios. The Decision-WM resulted in the best performance from $M=3-6$ with an $EER \leq 0.006$ for the normal test and $EER \leq 0.010$ for the leaked test. In previous studies, the Decision-WM has been preferred when the classification accuracy of individual matches vary, and it has also been suggested that uncorrelated modalities significantly improve authentication performance [77, 78]. For the current study, the weights employed are the classification accuracy of individual gestures, which are presented in [Appendix B](#). For $M=5-6$, the *EER* for both scenarios was 10^{-4} , indicating superior biometric performance even when the code was compromised. These findings suggest that with Decision-WM, Score-Norm, and a lower codelength ($M \leq 3$), the multicode EMG system provides an additional KW security to the existing

BM security. Furthermore, when the codelength is higher ($M > 3$), the BM security is greatly enhanced such that the KW security is redundant.

For the identification mode, the results demonstrated that the *FAR* of the multicode system remained similar for all the fusion schemes. The *FRR* base analysis showed that both the Score-Norm and Decision-WM had relatively better performance than all the fusion schemes. Detailed analysis suggested that while for a rank-1 system and using lower codelengths ($M \leq 3$), the Score-Norm had a significantly lower *FRR* than Decision-WM. However, for higher ranks ($K > 1$) and higher codelengths ($M > 3$), it was significantly different. The results showed that $FRR_{m,4}$ for Score-Norm reached a plateau of 10^{-4} for higher codelengths ($M > 1$). Score-Norm utilizes a w-score based normalization, which significantly outperforms the traditional min-to-max normalization. This finding is in accordance with a previous multimodal identification study that employed w-score based normalization to improve the recognition accuracy [122]. Rank-LR and Rank-Borda had a relatively poorer performance than the Score-Norm, with Rank-LR having the worst. This is in contrast to a previous multimodal biometric study where the Rank-LR had a better performance than the Rank-Borda scheme [91]. However, that study utilized facial, ear, and signature recognition, which was different from the nature of the gesture-based EMG signals. Therefore, the results demonstrated that based on the mean *FRR* values of all the fusion methods, Score-Norm had the best identification performance.

4.4.2 Parameter (Codelength and Rank) Analysis

As the rank parameter is only applicable to the identification mode, only the codelength was used to determine the best authentication performance. From our primary analysis, it was determined that for higher codelengths ($M > 2$), the Decision-WM outperformed all other fusion schemes in the normal-test and leaked-test scenarios of the authentication mode. However, the Score-Norm, showed comparable performance and at lower codelengths ($M \leq 2$), it was better than Decision-WM. It was clear that the multicode performance was significantly better than the unicode performance. Detailed analysis suggested that *EER* values as low as 0.005 (for normal test) and 0.007 (for the leaked test) were achieved using the 4-code EMG system. Therefore, the codelength of four was considered desirable for the multicode authentication mode.

For the identification mode, the codelength did affect the *FAR*; hence, only the rank was considered for determining the desirable *FAR*-based performance. In terms of the lowest *FRR* values, the Score-Norm fusion had a superior performance as follow: For a rank-1 system, the lowest *FRR* was obtained for a codelength of six ($M=6$, $FRR=0.001$). For the rank-2 system, the lowest *FRR* (10^{-4}) was obtained for a codelength of 4. For the rank-3 and rank-4 systems, the lowest *FRR* (10^{-4}) was obtained for three codelengths ($M=3$) and two ($M=2$), respectively. However,

Table 4.1: Recommended Parameter Values of Multimodal EMG Biometric System

Authentication	Fusion	Codelength	Current Study <i>EER</i>		NinaPro DB <i>EER</i>	
Normal Test	Decision- WM	Unicode	0.027		0.062	
		4	0.005		0.026	
		Unicode	0.037		0.064	
Leaked Test		4	0.007		0.022	
Identification		Codelength	Current Study		NinaPro DB	
			<i>FRR</i>	<i>FAR</i>	<i>FRR</i>	<i>FAR</i>
Rank K=1	Score- Norm	Unicode	0.165	0.009	0.120	0.001
		6	0.001	0.009	0.001	10^{-4}
Rank K=2		Unicode	0.062	0.047	0.050	0.017
		4	10^{-4}	0.047	10^{-4}	0.016
Rank K=3		Unicode	0.029	0.089	0.028	0.034
		3	10^{-4}	0.089	10^{-4}	0.033

EER: Equal Error Rate; *FRR*: False Rejection Rate *FAR*: False Acceptance Rate

Q1: 1th quartile (25%) Q3: 3rd quartile (75%)

the results demonstrated that the *FAR* values increased linearly with the rank of the system, irrespective of the codelength. This finding was expected, as a higher-ranked system grants access to a greater number of users, of which only one is the genuine user and the rest are impostors. For the rank-1-, rank-2-, and rank-3 system, the observed *FAR* with the Score-Norm fusion was 0.009, 0.047, and 0.089, respectively, which were lower than *FAR*=0.1. Because the recognition accuracy of identification mode deteriorates with a greater number of users, a rank-1-, rank-2-, or rank-3 system can be chosen accordingly, and their recommended codelengths are six, four, and three, respectively. Table 4.1 summarizes the performance metrics using recommended parameter values and the best fusion schemes for a unicode and multicode EMG biometric system. Additionally, for the recommended values of codelength and rank, the analysis was repeated for the combined NinaPro DB2 and DB7 database consisting of 60 participants and performing 17 hand and finger gestures [4, 76]. It was found that the performance was comparable to the data from the experimental data set obtained in the present study, for both the authentication and the identification scenarios. Compared to the present study, the unicode *EER* was higher for the normal test and leaked test, which were reduced from 0.063 and 0.064 to 0.026 and 0.022, respectively, for codelength at four (M=4). For the identification scenario, the performance was better than the present study (see Table 4.1), thus corroborating our findings.

The multicode EMG biometric performance analyzed in this study was either comparable or

slightly better than that of previous studies that utilized EMG-Biometrics [44, 64, 65, 121]. The unicode scenario in the present study was similar to our previous findings while using a four channel EMG and TD extraction method. [105]. Furthermore, the *FRR* of rank-1 was 0.03. The higher unicode error rates in this study (Table 4.1) can be attributed to the reduced dataset used for training (two trials for training compared to the six trials for the previous study). This reduction in data was done to match a real-world scenario and reduce the training burden. In addition, the present study utilizes a mean value of 100 randomly generated code sequences as compared to the previous study [105]. Another unicode study that utilized eight electrodes on the forearm while using an FDT feature set reported a unicode authentication *EER* of 0.035 and a *FRR* for rank-1 of 0.097 [44]. However, the multicode system error rates (see Table 4.1) were lower than those of unicode studies [4, 44]. Another study utilized a fixed codelength ($M=8$) for the finger gestures as a neuromuscular password [64]. The normal-test *EER* was 0.012 and the leaked-test *EER* was 0.1496. Another identification study with a similar setup involving a fixed codelength ($M=8$) reported a recognition error rate of 0.142 [62]. A different study allowed users to choose three out of the four gestures as a code for personal authentication [121]. They reported a authentication *FAR* of 0.002 and an *FRR* of 0.2. The lower error rates of the present study in comparison to the previous studies might be because all the studies utilized a pre-match level fusion where the EMG signals were combined at the source level, and thereafter feature extraction and template matching were performed. This study demonstrated the use of such multiple gestures as an authentication code as well as a biometric trait by employing post-match fusion at the score-, rank-, and decision level. Post-match fusion not only results in a better performance but allows the user a customizable system where they can set access codes of any desired codelength. This facilitates the realization of the true potential of the dual security (KW and BM) of gesture-controlled forearm EMG signals. The concept of customizable codes in the current study is similar to the cancelability and cross-application biometrics proposed in [62], with the difference in fusion level, the post-matcher level vs. the pre-matcher level. In addition, the identification performance, which was hypothesized to be worse than the authentication mode, was found to be comparable to the authentication performance using a multicode system. For instance, a low recognition error ($FRR=10^{-4}$) can be achieved using a code sequence of four gestures in a rank-2 system. Both the authentication and identification error rates are comparable to traditional biometric traits such as fingerprint, DNA, iris, and facial recognition.

4.4.3 Limitations and Future Work

Although promising results were obtained for the multicode EMG biometric system, one limitation was that the study involved data from 24 participants. This number was similar to other EMG-Biometrics studies [62, 64], however, the subject pool size is small compared to research

involving conventional biometric traits and further studies involving a larger database size is important to corroborate results. In the current study, the Rank-LR and Rank-Borda performed poorly compared to other methods. However, with a larger subject pool, they may provide improved results. While this might be of future interest, it was important to study the relationship of the population size on the EMG biometric system performance. Another key limitation was that the present study involved only one experimental session for data recording, which was partly because of the COVID-19 constraints currently in place in the University of Waterloo, Canada. To better address robustness issues such as electrode placement, ability to follow instructions and background noises, multi-day trials should be performed. For this study, the order (sequence) of the codes was not considered, so as to lay emphasis on the comparison of the fusion schemes. It was reasonable to expect that even better performance would be obtained when the order of the codes is included, which will be investigated in a future study.

4.5 Chapter Conclusions

The results presented in the current study demonstrated that the proposed multicode EMG biometric system has an overall superior performance compared to the unicode system in both authentication and identification applications. Specifically, the Decision-WM fusion for the authentication mode and Score-Norm fusion for the identification mode had the best performance, which suggests them as recommended fusion schemes for the two applications, respectively. It was found that the recommended codelength for authentication mode was four. In addition, the biometric security for such a system was highly improved, rendering KW security redundant. A very low *FRR* (0.01%) was obtained with the rank-1-, rank-2-, and rank-3 identification for a codelength of 6, 4, and 3, respectively. For such a system, the rank played a dominant role in determining the system's *FAR*, as there seemed to be a linear relationship between rank and *FAR*. The flexibility of choosing desirable parameter values allowed the system's performance to be adjusted to the population size. Therefore, an improved multicode EMG biometric system with dual security can be designed using specific fusion schemes while using the recommended parameter values of code length and rank.

Chapter 5

Multi-day Dataset of Forearm and Wrist Electromyogram for Hand Gesture Recognition and Biometrics

This chapter was published in the *Nature's Scientific Data* vol 9 as Pradhan, A., He, J. & Jiang, N., “Multi-day dataset of forearm and wrist electromyogram for hand gesture recognition and biometrics”, Article 733 (2022). This data descriptor details the data collection methodology of a multi-day and large-sample EMG dataset [107].

This is a data descriptor article that provides extensive details on the GRABMyo data acquisition setup and protocol, followed by the technical validation of the dataset.

5.1 Chapter Introduction

Surface electromyography (EMG) signals have been used for advanced prosthetics control, hand-gesture recognition (HGR), and more recently as a novel biometric trait. For these EMG-based applications, the translation from laboratory research settings to real-life scenarios suffers from two major limitations: (1) a small subject pool, and (2) single-session data recordings, both of which prevent acceptable generalization ability. In this longitudinal database, forearm, and wrist EMG data were collected from 43 participants over three different days with long separation (Days 1, 8, and 29) while they performed static hand/wrist gestures. The objective of this research is to provide a comprehensive dataset for the development of robust machine learning algorithms of EMG signal processing, for both HGR and biometric applications. The high quality of the

current dataset was demonstrated by comparing with the Ninapro dataset. The usability of the dataset for both HGR and biometric applications was presented. Among other applications, the dataset can also be used for developing electrode-shift invariant generalized models, which can further bolster the development of wristband and forearm-bracelet sensors.

5.2 Background & Summary

Recent advances in machine learning techniques have enabled applications of hand gesture recognition (HGR) using surface EMG signals. This has further boosted the development of advanced prosthesis control systems for rehabilitation of upper limb amputees [55], and recent industrial applications have emerged using HGR for human-machine interactions in industrial [115] and consumer applications scenarios [142], [27]. For these applications, the EMG signal features are extracted and used as inputs for various machine learning techniques such as linear discriminant analysis (LDA) [28], support vector machines (SVM) [135] etc. for detecting hand gestures. More recently, advanced techniques such as deep neural networks (DNN) have achieved highly accurate classification performance even with simple architectures [19], [18].

However, extensive investigation has demonstrated EMG-based hand HGR has poor cross-user transference performance [10], suggesting an one-size-fits-all model for all users is still elusive. This indicates that EMG signals might contain biometric information as they contain individual differences. This has provided motivation for investigating the potential of EMG as a biometric trait. Combined with the HGR property, it enables the user to set user-defined gestures as a password for enhanced security, which is not possible with other bio-signals such as electroencephalogram (EEG) and electrocardiograph (ECG). Our recent studies have provided a framework for the fusion of these codes and to facilitate such a dual-mode (password and biometrics) security [106].

Although, high performance of both HGR and biometric models has been previously reported, there always exists a gap between the real-world conditions and the laboratory settings, under which most of the current HGR and biometric research have been conducted conditions. It has been established in the literature, that in a multi-session protocol spreading across days, non-stationary factors including electrode shifts, skin and physical conditions will seriously affect the performance of an EMG processing system [5]. As there are numerous such factors, experimentally controlling each of them would increase the number of trial repetitions exponentially and hence can be arduous. Nevertheless, a multi-day dataset with a sufficiently large subject pool is warranted for validating the effectiveness of EMG applications such as HGR and biometrics.

Some open-access databases of multi-day EMG recordings of forearm muscles are publicly available [2, 17, 59, 69, 139]. Two databases with large subject pool (>40) involved two days

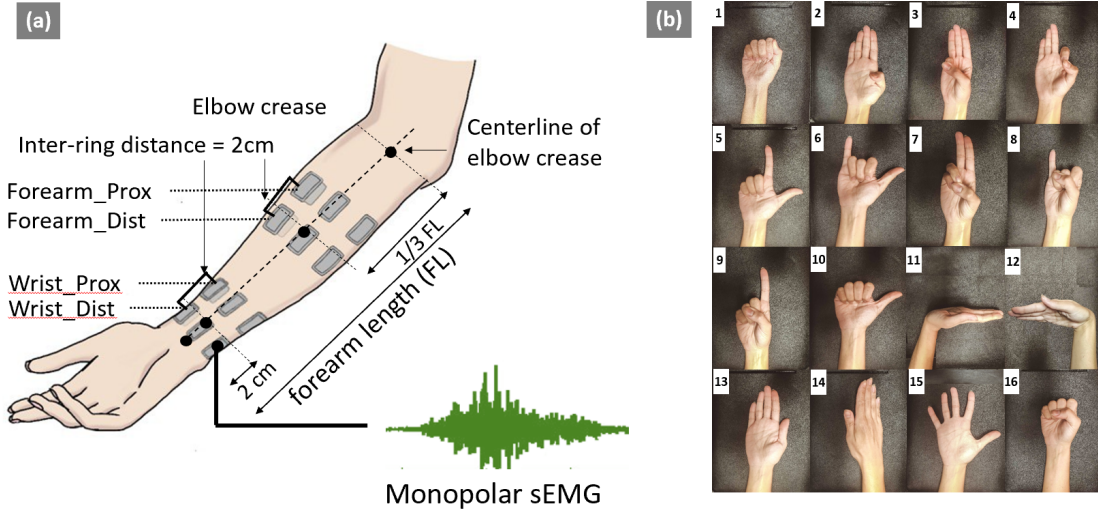


Figure 5.1: **Electrode Positions and Gesture list.** Fig 1A (left) shows the electrode positions used in the study. There are two groups of electrodes: forearm and wrist. Each group comprises of two rings (proximal and distal). Monopolar EMG was recorded from a total of 28 channels (16 Forearm and 12 wrist). Fig 1B (right) shows the Sixteen gesture classes investigated in the study (numbered 1 – 16): lateral prehension (LP), thumb adduction (TA), thumb and little finger opposition (TLFO), thumb and index finger opposition (TIFO), thumb and little finger extension (TLFE), thumb and index finger extension (TIFE), index and middle finger extension (IMFE), little finger extension (LFE), index finger extension (IFE), thumb extension (TE), wrist flexion (WF), wrist extension (WE), forearm supination (FS), forearm pronation (FP), hand open (HO), and hand close (HC). The REST gesture was also collected in each trial repetition.

of data collection [69, 139]. While one of them has as low as six channels [15], some others utilized a high-density (HD) EMG setup [2, 17, 59]. Only three studies involved more than two of data collection days, but the number of subjects was smaller (<11) [2, 31, 36]. Only one study with three days of data recording, had a sample size of 20 subject, with only signals from the forearm [23]. To explore the robustness and accuracy of HGR and biometrics, it is imperative to have a database with larger subject pool sizes, recorded across multiple days and comprising numerous gestures.

In the current study, we present an open-access named Gesture Recognition and Biometrics electroMyogram (GrabMyo) Dataset [56, 57]. GrabMyo consists of 43 participants (subsequently termed as users), three sessions in three separate days (subsequently termed as sessions) of data collection, 16 hand and finger gestures each with seven repetitions (subsequently termed as tri-

als). To our best knowledge, the presented dataset is the largest EMG dataset in terms of the total number of recording sessions (43 users $\times$ 3 days = 129 recording sessions). A unique feature of GrabMyo is that the EMG signals were recorded from both forearm and wrist positions. A graphical representation of electrode positions and the list of gestures investigated in the study is shown in [Figure 5.1](#) A and B. The sampling frequency of the recorded signals was selected as 2048 Hz. To obtain generalizable data, special effort was taken such as 1) electrode positioning protocol for each session, 2) normal level force instruction, 3) rest duration for avoiding fatigue, 4) un-uniform interval between data collection sessions (Days 1, 8, and 29) and 5) data collection from healthy users (subjects with a single session of sickness have been eliminated from the study). These considerations are explained in greater detail in the Methods section. The dataset provides a valuable resource for EMG-based HGR and biometrics research, particularly for improving algorithms' robustness in a multi-day scenario and cross-user generalization ability.

5.3 Methods

Subjects and Ethical Requirements: We recruited 43 healthy participants (23 M, 20 F) for the study spanning three different days: day 1, day 8 and day 29. The participants were students and staffs from the University of Waterloo. The average age was 26.35 ± 2.89 , and the average forearm length (measured from the styloid process on the wrist to the olecranon on the elbow) was 25.15 ± 1.74 cm. More details about the dataset and the participant characteristics are reported in [Table 5.1](#). Individuals with any existing muscle pain, skin allergies, and those who were unable to complete the three sessions due to any unprecedented circumstances were excluded from the study. For the third session (day 29), a range of 3 days was provided for some participants who couldn't make it to the specific appointment. Before the enrolment, the participants were provided an oral and written explanation of the procedures and signed an informed consent form. They were informed that they could withdraw from the study at any point. The experiments were conducted following the Declaration of Helsinki and the research protocol was approved by the Office of Research Ethics of the University of Waterloo (ORE# 31346). The participant demographics are provided in [Appendix C](#).

Acquisition Setup: The experimental setup consisted of a PC and a monitor mounted on a desk, 0.75 m in front of a height-adjustable chair. The EMGUSB2+ (OT Bioelettronica, Italy), a commercial amplifier, was used for acquiring the EMG signals. The gain of the device was set to 500, and the sampling rate was set to 2048 Hz. Pre-gelled skin-adhesive monopolar EMG electrodes (AM-N00S/E, Ambu, Denmark) were used.

Prior to the experiment, the user's forearm length is measured as the distance between the

Table 5.1: Database Summary

Participant Characteristics	Values
# Males (# Females)	23 (20)
# Right-handed (#Left)	38 (5)
Age (years)	26.35 ± 2.89
Forearm length (cm)	25.15 ± 1.74
Forearm circumference (cm)	24.10 ± 2.27
Wrist circumference (cm)	16.18 ± 1.21

olecranon process and the ulnar styloid process. The forearm circumference is measured at one-third of the forearm length from the olecranon process. The wrist circumference is measured at 2 cm away from the ulnar styloid process. Prior to electrode placement, the skin surface was shaved to remove hairs, cleaned with an alcohol swab, and abraded with a paper towel. For the forearm electrode placement, sixteen EMG electrodes were placed in the form of two rings, each consisting of eight electrodes equally spaced around the forearm, forming eight bipolar pairs. The center-to-center distance between the two forearm-rings was maintained at 2 cm. For the wrist electrode setup, twelve monopolar EMG electrodes of the same type as the forearm rings were placed in the form of two rings, each consisting of six electrodes equally spaced around the wrist and forming six bipolar pairs. The center-to-center distance between the two wrist-rings was maintained at 2 cms, the same as the forearm setup. Therefore, a total of 28 monopolar EMG electrodes were used for each session, which forms four electrode rings: proximal and distal rings for the forearm and wrist. A detailed pictorial representation is provided in [Figure 5.1 1A](#). To maintain consistency of the positions of the electrodes across all participants, the first electrode in each ring (total rings = 4) was anatomically positioned on the center-line of the elbow crease as shown in [Figure 5.1 1A](#) [44, 105]. Although a standardized electrode placement protocol was followed for each session, no marks were left on the participant’s skin. This intentionally induced uncertainty in the exact electrode positions corresponds to the real-life scenario where it is not realistic, if possible at all, to place an electrode band at the same position across multiple sessions. The physical electrode locations are shown in [Appendix D](#).

Acquisition Protocol: For each experimental session, following the acquisition setup, the participant is seated comfortably on the chair with both their upper limbs in a resting position. Visual instructions for performing the gestures were provided on the computer screen placed in front of the participants. The participants were instructed to perform the gestures at a normal force level, or similar to how they would normally do it during daily activities. For an estimate of “normal force”, the participants were asked to perform multiple trial contractions at three self-defined force levels: soft, hard and medium contractions, where a medium level corresponds to

the contraction of the normal force. The following 16 hand and wrist gestures were included in the current study (presented in Figure 5.1 1B): Lateral prehension (LP), thumb adduction (TA), thumb and little finger opposition (TLFO), thumb and index finger opposition (TIFO), thumb and little finger extension (TLFE), thumb and index finger extension (TIFE), index and middle finger extension (IMFE), little finger extension (LFE), index finger extension (IFE), thumb extension (TE), wrist flexion (WF), wrist extension (WE), forearm supination (FS), forearm pronation (FP), hand open (HO), and hand close (HC). The order of the 16 gestures was randomized and a resting (REST) trial was collected after all 16 gestures were performed once. A ten-seconds relaxing period was provided between each trial. One continuous data acquisition of 17 gestures (including the REST) is called one run. First, a trial run was performed to ensure that the participants understood the experimental protocol. Following the trial run, seven runs were recorded for each user, resulting in 119 contractions (17 x 7). Any accidental gesture or no-activity/delayed-activity was noted and the respective gesture's replacement contraction were performed after each run. The user could also request additional rest when he/she felt necessary. The entire session was repeated on day 8 (after 1 week) and day 29 (after 1 month).

Signal Processing: The EMG signals were bandpass filtered between 10 Hz and 500 Hz using a fourth-order Butterworth filter. A notch filter of 60 Hz was employed to remove the powerline noise that might have affected the signal recording.

5.4 Data Records

Data records [56, 57] presented in this section and accompanying description files are available online in PhysioNet (<https://physionet.org/about/database/>) and IEEE Dataport (<https://ieee-dataport.org/datasets>). The database consists of 43 participants, three sessions in three separate days (subsequently termed as sessions) of data collection, and 17 gestures (including REST) each with seven repetitions. All the EMG recordings are 5 seconds in duration, collected from 28 channels (16 forearm and 12 wrist), and sampled at 2048 Hz.

The components of the released repositories are described in detail (Table 5.2) and are organized as follows: waveform data and additional files.

Waveform Data: For the PhysioNet Database, the EMG recordings are provided in the waveform database (WFDB) format, which is considered the most widely used medium for storing physiological signals and waveform data [37]. There exist numerous open-source WFDB libraries for commonly used analysis tools using MATLAB and Python. The WFDB allows a structured way of storing the EMG recording in the form of a tuple of two files: a dat-file containing the binary raw data and a corresponding header file with the same name and a hea-extension.

Table 5.2: Multi-Day EMG Database. MATLAB Codes: File Reading and Signal Processing

Repository	Folders	Sub-Folders	Files	Description
PhysioNet	Session 1	sessioni_subjectj	sessioni_subjectj_gesturek_triall.dat	$i \in [1, 2, 3]$ represents the session(day) index
	Session 2		sessioni_subjectj_gesture k _triall.he (dat file contains 10240×32 hexadecimal values.	$j \in [1, 2, \dots 43]$ represents the subject index
	Session 3		hea file contains signal information such as sampling frequency, units and gain)	$k \in [1, 2, \dots 17]$ represents the gesture index
	Session 3			$l \in [1, 2, \dots 7]$ represents the trial index
IEEE Dataport	Session 1	subjecti_sessionj	sessioni_subjectj.mat (.mat file contains 17×7 cell matrix. each cell contains 10240×32 numeric array)	$i \in [1, 2, 3]$ represents the session(day) index
	Session 2			$j \in [1, 2, \dots 43]$ represents the
	Session 3			subject index
Additional Files				
PhysioNet	< N/A >	< N/A >	readme.txt, Subject-info.csv, MotionSequence.txt, GestureList.jpg, DeviceInfo.pdf, Electrodelocation.pdf	readme.txt general information of the dataset and reading data files.
				MotionSequence.txt provides gesture definitions and their sequence
				GestureList.jpg provides pictorial representation of hand gestures.
IEEE Dataport	< N/A >	< N/A >	Subject-info.csv, MotionSequence.txt, GestureList.jpg	DeviceInfo.pdf describes the configuration of acquisition device.
				ElectrodeLocation.pdf illustrates the positioning of electrode sensors.
				Subject-info.cs v contains participant anthropometric information

The hea-extension contains all the signal-specific metadata such as the channel names, sampling frequency, and the scaling factor for converting the signal to physical units (in mVs).

For the IEEE database, the signal files are saved as *.mat* files. The mat file extension allows users a convenient approach for obtaining the EMG recordings in the physical unit format (additional conversion is not necessary). Additionally, the mat files allow data to be organized as cells where multiple gestures and repetitions from a single session can be presented together.

Additional Files: The additional files provide supplementary information such as the participant anthropometrics, electrode positioning guidelines, device configuration, the sequence of gestures, and their descriptions. MATLAB codes are also provided for reading the data files and subsequent feature extraction as explained in the Code Availability section.

5.5 Technical Validation

A specific trial of an individual user case was chosen at random and the amplitude and frequency components were analyzed as shown in [Figure 5.2](#). It was observed that all 28 channels of forearm and wrist possessed equal amplitude and power spectrum distribution. In the following sections, signal processing techniques are discussed for analysis and technical validation of the EMG signals. As well, the original references where the methods were described in depth are provided. For comparison, the validation results are provided along with those for the widely used Ninapro DB2: Exercise B comprising 40 users and 17 gestures.

Signal-to-Noise Ratio (SNR) The spectral properties of each EMG signal recording were analysed and the *SNR* (in dB) was measured as the ratio between the power of the signal to the power of the noise [8, 129]. As the types of the artifacts were unknown in our case, the power of the noise was estimated as the power of EMG recordings during the rest trial. The average *SNR* of all the signals (14.565 ± 6.385 dB) was in range with the *SNR* values suggested for the wrist and forearm [8] and they are consistently higher than those from the Ninapro database (8.001 ± 4.051 dB) (see [Figure 5.3 A](#)).

Correlation coefficient of Normality (CCN) The *CCN* was measured to analyze the amplitude distribution. For a static contraction with moderate force, the EMG can be modeled as a filtered, random, white Gaussian noise process [128]. It has been suggested that a test of normality can provide a measure of biosignal quality, where a signal amplitude with a non-normal distribution would be considered contaminated. A Gaussian distribution with equal mean and variance to that of the recording is generated [33]. The *CCN* is defined as the Pearson correlation coefficient between the histogram bin values of the EMG recording and the normal density

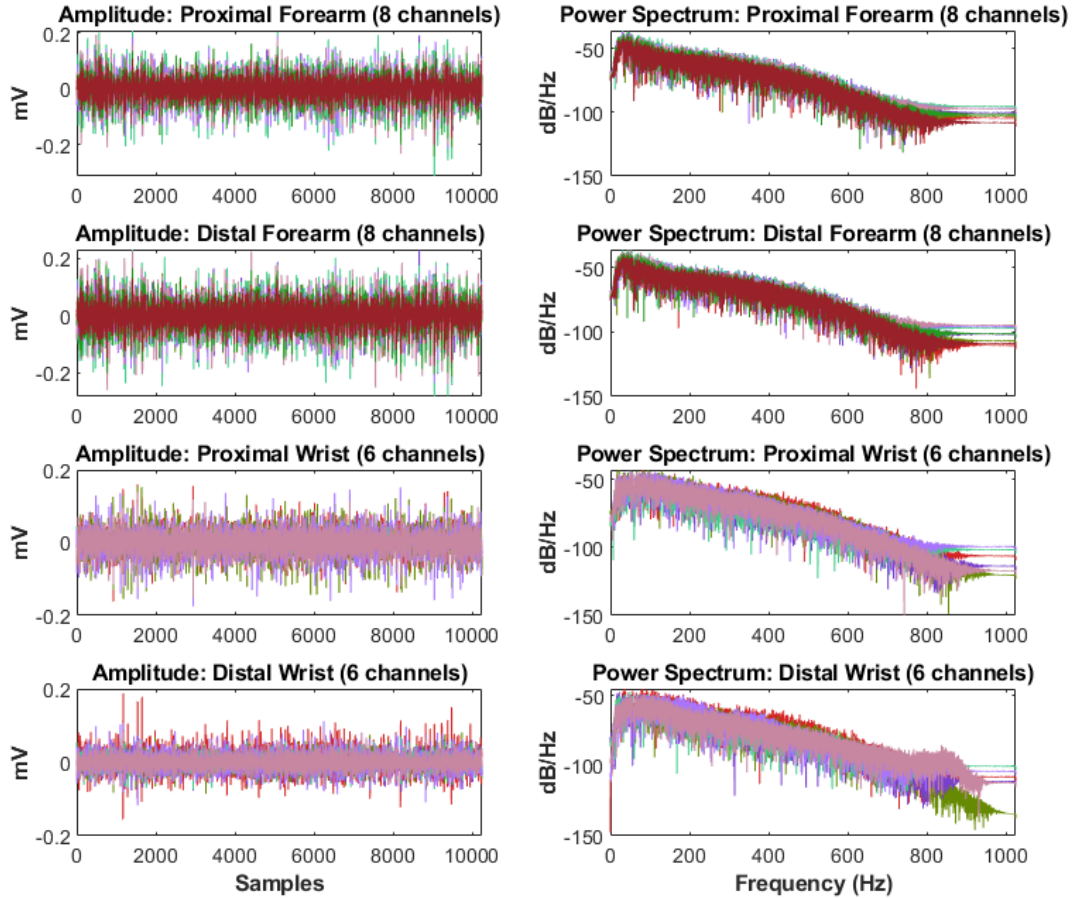


Figure 5.2: **Representative amplitude and power spectrum plots.** Left Column: The amplitude values (in mV) for the four electrode rings: proximal and distal forearm and wrist (from top to bottom). Right Column: The fast fourier transform (FFT) power spectrum (in dB/Hz) for the four electrode rings: proximal and distal forearm and wrist (from top to bottom). The forearm rings constitute 8 channels each and the wrist rings constitute 6 channels each (shown in different colors).

function value for the corresponding bins. A value close to 1 is considered a normal distribution. It was observed that the *CCN* of all the signals (0.975 ± 0.041) was close to 1 which was consistently higher than those from the Ninapro database (0.848 ± 0.075) (see [Figure 5.3 B](#)).

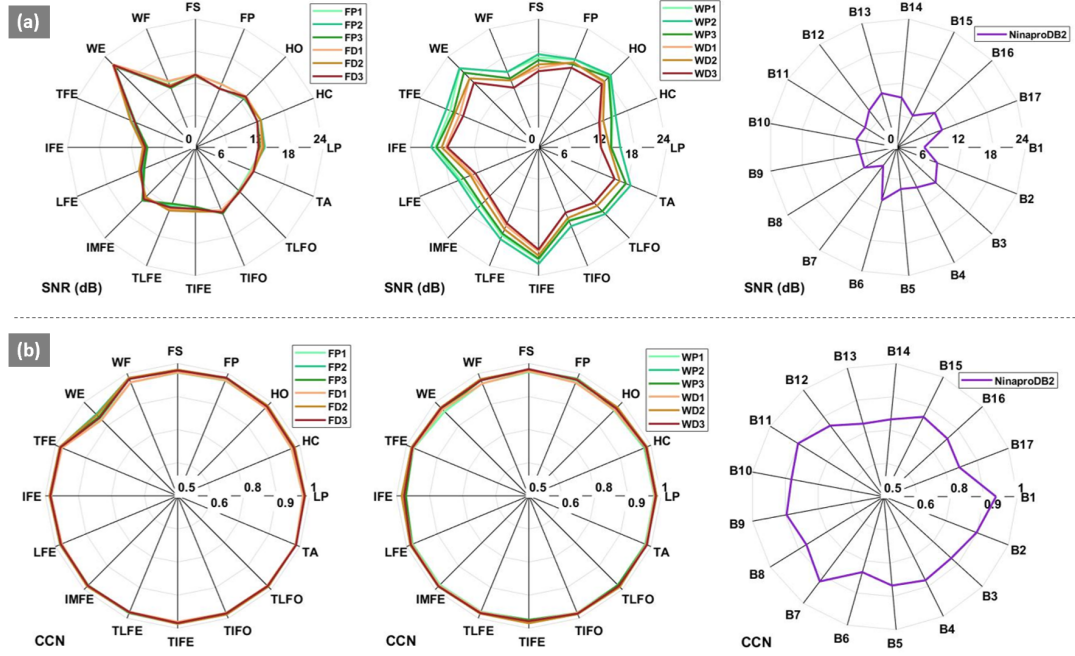


Figure 5.3: **SNR and CCN comparison plots.** Panel A, i.e. the top row shows *SNR* values (dB), with each spoke representing one motion. Left: the forearm proximal and distal electrode setup for session 1-3 (added as suffix FP1, FP2,...,FD3); middle: wrist proximal and distal electrode setup for session 1-3 (added as suffix WP1,WP2,...,WD3); and (right) the Ninapro signals collected from the forearm for a single day. Panel B, i.e. the bottom row shows the *CCN* values for the forearm (left), wrist (middle) and Ninapro data (right) in a similar manner.

Standard HGR and Biometric Analysis

Prior to HGR and Biometric evaluation, the signals were first processed and then features were extracted as described below. The forearm rings (eight channels) and the wrist rings (six channels), the monopolar EMG signals were first re-referenced by a common average procedure. The processed signals were then segmented into 200ms-width windows, with a 150 ms overlap. Each window was then processed using Hudgins's time-domain (TD) feature extraction [52]. Time-domain features (mean absolute value, zero crossing, slope sign changes, and waveform length) were extracted from filtered data [105]. Therefore, for a forearm setup, the feature vector consisted of $8 \times 4 = 32$ features, while the wrist setup consisted of $6 \times 4 = 24$ features.

For the biometric analysis, a matching score, commonly the Mahalanobis distance, is used to assess if it's a match (access granted) or no match (access denied) [44,105,106]. To maintain con-

Table 5.3: Mean ($\pm$ STD) HGR and Biometric performance.

Electrode Setup	HGR Evaluation (<i>AUC</i>)	Biometric Evaluation (<i>EER</i>)
Forearm	0.948($\pm$ 0.018)	0.028($\pm$ 0.007)
Wrist	0.941($\pm$ 0.021)	0.038($\pm$ 0.006)
Ninapro (Forearm)	0.875($\pm$ 0.034)	0.038($\pm$ 0.013)

sistency in the HGR and biometric analyses, a Mahalanobis distance classifier was implemented for both analyses. For a given feature vector sample p (the input), its Mahalanobis distance $S_{i,j}$, with the i th gesture and the j th user, was defined as

$$S_{i,j}(p) = \sqrt{(p - \mu_{i,j}) \Sigma_{i,j}^{-1} (p - \mu_{i,j})^T} \quad (5.1)$$

where $\mu_{i,j}$ is the centroid of the i th gesture class and the j th user and $\Sigma_{i,j}$ is the covariance matrix for the specific gesture and user class. Both the parameters are calculated from the system training data and the sample p is from the system testing data. The leave-one-out (LOO) cross-validation scheme was used, where six trials were used for training and one trial for testing. Figure 5.3A, Figure 5.3B, and Table 5.3 demonstrate the results of the technical validation as described in the following sections.

HGR evaluation: In this study, the HGR analysis was performed in a user-specific scheme. For a particular user and a particular gesture, the true class consisted of the feature vectors from the target gesture of the user and the false class consisted of the feature vectors from the remaining 15 gestures of that user (the rest gesture was excluded from the HGR analysis to maintain consistency with biometric analysis). Similarly, for the Ninapro database, the true class consisted of feature vectors from the target gesture of a specific user and the false class consisted of the feature vectors of the remaining 16 gestures. For the performance analysis a receiver operating characteristic (*ROC*), where the true positive rate (*sensitivity*) was plotted against the false positive rate ($1 - \textit{specificity}$) by varying the threshold distance of correct gesture prediction. The true positive rate or sensitivity represents the probability of detecting a correct gesture, while the false positive rate is the probability of detecting an incorrect gesture. The area under the curve (*AUC*) is calculated from the *ROC* curve [7]. The *ROC* curve and the *AUC* values for all the users, days, and gestures are averaged and reported separately for the forearm and wrist electrode positions. Figure 5.4 A shows the *ROC* plots for HGR analysis. It was observed that the *AUC* for the forearm was 0.948 ($\pm$ 0.018) and for the wrist was 0.941 ($\pm$ 0.021). Both the values were comparatively higher than the corresponding *AUC* value of 0.875 ($\pm$ 0.034) for the NinaPro data.

Biometric evaluation: In this study, the authentication mode of biometrics was used to demonstrate the feasibility of multi-day EMG-based biometrics. In this mode, the true user's identity and the corresponding gesture are known to the system. As such, for a specific user, the true class consisted of the feature vectors from the target gesture (*e.g.* the passcode) of that user and the false class consisted of the feature vectors of the remaining 42 users for that gesture. Similarly, for the Ninapro database, the true class consisted of feature vectors from the target gesture of a specific user and the false class consisted of the feature vectors from the remaining 39 users for that gesture. For the performance analysis a receiver operating characteristic (*ROC*), where the true positive rate (sensitivity) was plotted against the false positive rate ($1 - \text{specificity}$) by varying the threshold distance of correct biometric matching. The true positive rate or sensitivity represents the probability of detecting a correct hand gesture, while the false positive rate is the probability of detecting an incorrect hand gesture. The equal error rate (*EER*), is obtained from the *ROC* curve, where the false positive rate is equal to the false negative rate ($1 - \text{sensitivity}$) [7]. The *ROC* curve and the *EER* values for all the users, days, and gestures are averaged and reported separately for the forearm and wrist electrode positions. Figure 5.4 B shows the *ROC* plots for Biometric analysis. It was observed that the *EER* for the forearm was $0.028 (\pm 0.007)$ and for the wrist was $0.038 (\pm 0.006)$. Both the values were comparatively lower than the corresponding *EER* value of $0.038 (\pm 0.013)$ for the NinaPro data.

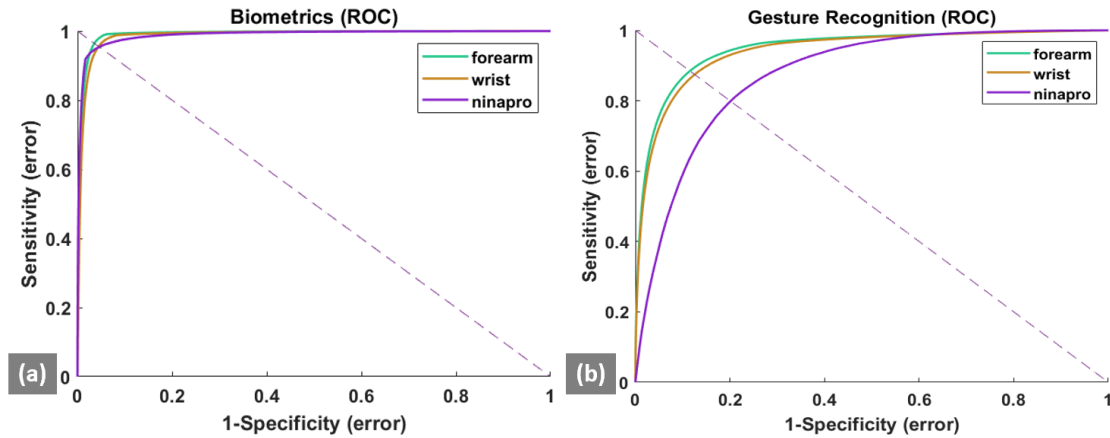


Figure 5.4: **ROC curves for HGR and Biometric validation.** The *ROC* curves are plotted for the HGR evaluation (A) and Biometric Evaluation (B) for the forearm and wrist data averaged for the multiple users ($N=43$), sessions ($N=3$) as well as for the Ninapro Data ($\#users = 40$, $\#sessions = 1$). The dotted line represents the point of intersection on the *ROC* curve from which the *EER* value is obtained.

5.6 Usage Notes

As evident from the title, the GRABMyo Dataset has two major applications: HGR using machine learning approaches and EMG-based biometrics. The true potential of GRABMyo are three-fold: 1) large subject-pool, 2) multi-day session, and 3) both wrist and forearm channels. Through the Signal Analysis section, we will explain processing EMG recordings and the cross-validation approach for obtaining reliable benchmark results using machine learning algorithms. Following this, the Future Direction lists novel research strategies for HGR and Biometric analysis which could be bolstered using the large multiday dataset.

Signal Analysis The MATLAB code `fileread.m` allows the users to read data files along with their signal-specific metadata. After conversion to a numeric format, the data is structured as matrixes with timepoints as rows and monopolar EMG channels as columns. The channel names provided can be used for separate forearm and wrist electrode rings and also for forming bipolar pairs between them [132]. The processed data is then windowed, and features can be extracted using multiple techniques [97]. A widely used frequency division technique [109] is provided as a sample feature extraction in `feature_extraction.m`. The gesture labels are provided by `Motion-Sequence.txt` which allows the extracted features to be matched to the specific gestures and their respective users. For HGR and biometric applications, the output classes are the gestures and users, respectively which can be used for developing machine learning and deep learning architectures. The data provides seven trial repetitions, hence seven-fold cross-validation (six-folds for training and one-fold for testing) can be used to simulate a near-practical scenario, where the testing data is recorded separately from the training data. A more robust analysis *i.e.*, the multi-day analysis with a 3-fold cross-validation can be employed, where subsets of data from two days are used for training and the data from the remaining day is used for testing.

Future Directions Some future research directions are:

Improving biometric authentication: One unique advantage of the EMG biometric trait, in particular for authentication applications, is the combination of user-specific biometrics with user-defined gestures as passcodes, the latter of which is not possible with other bio-signals such as electroencephalogram (EEG) and electrocardiogram (ECG). A multi-code EMG-Biometrics framework can be used to combine the gestures and improve authentication performance and security [106].

Biometric identification Another major biometric application is the identification mode where the system predicts the identity of the presenting user by finding the closest match. The identification is a more error-prone application as the system makes N comparisons, where N is the number of users. Therefore, the factors affecting system performance such as multiple days and sample size of the database need to be investigated for real-life applications.

Subject independent gesture recognition Extensive research on EMG has been performed on gesture recognition with application in rehabilitation using prosthetic and orthotic devices, home application for assisting daily activities, virtual environment control, and sign language recognition [14, 27, 142]. Recent studies have suggested deep learning techniques for cross-user calibration-free which trains generalized models using the population data, and hence reduces the training burden of the user [10, 21, 24]. The presented large-sample dataset can provide resources for such calibration-free models.

Electrode shift-invariant techniques One of the significant factors affecting the cross-day EMG performance is the shift in the electrode positions. It is impossible to fix the location of armband electrodes on the forearm and wrist for daily-wear use. These variations affect the performance of both the EMG-based biometric and gesture recognition applications. Some techniques such as classification model adaptation [72, 140] and feature space transformation using transfer learning [21, 96, 112] have been suggested to address the electrode shift variations. These techniques could be further investigated to potentially improve biometrics and HGR performance.

5.7 Code Availability

The custom codes used for reading the signals of the database was created in MATLAB R2017b and is freely accessible at Physionet and IEEE Dataport [56, 57]. To implement the codes, the users will need a MATLAB License

1. A readme file (readme.txt) with instructions about how to run the code in a 2017b or higher MATLAB version.
2. A Matlab script (fileread.m) with a simple example about how to read WFDB files and convert them to .mat format.
3. A text file (MotionSequence.txt) which provides the gesture sequence, and thus can be used to assign class labels to input data.
4. A Matlab script (feature_extraction.m) allows a simple example to extract frequency features using featiDFTI.m and segmentEMG.m functions.
5. A Matlab function (featiDFTI.m) for generating frequency division technique features from EMG Data.
6. A Matlab script (segmentEMG.m) for implementing windowing of EMG Data.

Chapter 6

Multi-day Analysis of Wrist Electromyogram-based Biometrics for Authentication and Personal Identification

This chapter was published in the *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 5 as A. Pradhan, J. He, H. Lee and N. Jiang, “Multi-Day Analysis of Wrist Electromyogram-Based Biometrics for Authentication and Personal Identification,” article no. 4, pp. 553-565, Oct. 2023 [108]. This chapter investigates the multi-day performance of EMG-Biometrics while utilizing the GRABMyo dataset.

6.1 Chapter Introduction

Recently, EMG has been proposed for addressing some key limitations of current biometrics. Wrist-worn wearable sensors can provide a non-invasive method for acquiring EMG signals for gesture recognition or biometric applications. Previous studies have reported a high biometric performance for authentication and identification application [9, 30, 44, 60–62, 64, 75, 83, 86, 87, 101, 105, 121, 145]. EMG signals contain individuals’ information and can facilitate multi-length codes or passwords (for example, by performing a combination of hand gestures) [106]. However, current EMG-based biometric research has two critical limitations: small subject-pool for analysis and limited to single-session datasets [105, 106]. One of the most important features of a viable biometric trait is its longitudinal robustness across multi-session and multi-day. A few studies with a small subject pool (<22), analyzed data from a two-day protocol [61, 62]. Only one study had five subjects with a four-day data collection protocol [117].

It has been established in the EMG processing literature, that in a multi-session protocol spreading across days, non-stationary factors including electrode shifts, sweat, dry skin, and physical conditions will affect the accuracy and consistency of the EMG processing system [5]. Therefore, the multi-day performance with a sufficiently large subject pool is a crucial step for validating the effectiveness of EMG as a biometric trait. Furthermore, in a more practical identification mode scenario, intruders (unregistered users) might claim access to a biometric system [60]. An intruder analysis should be further performed, where an initial threshold-based grant/reject is employed to check for the authenticity of a claiming user and then followed by individual recognition.

For the scope of this paper, the multi-day biometric analysis was performed on the wrist-worn EMG over three sessions over the span of 30 days. The biometric benchmarks of both authentication and identification were presented. Additionally, the intruder analysis for the identification mode was investigated.

6.2 Methods

6.2.1 Participants

We recruited 43 healthy participants (23 M, 20 F) for the study. The average age was 26.4 ± 2.89 , and the average forearm length (measured from the styloid process on the wrist to the olecranon on the elbow) was 25.2 ± 1.74 cm. The average wrist circumference (measured at a distance of one cm away from the ulnar styloid process) was 16.2 ± 1.21 cm. Before the experiment, the participants were informed of the procedures and signed an informed consent form. The experiments were conducted following the Declaration of Helsinki and the research protocol was approved by the Office of Research Ethics of the University of Waterloo (ORE# 31346).

6.2.2 Acquisition Setup

The experimental setup consisted of a PC and a monitor mounted on a desk, 0.75 m in front of a height-adjustable chair. The EMGUSB2+ (OT Bioelettronica, Italy), a bio-signal commercial amplifier, was used for acquiring the EMG signals. The signals were bandpass filtered between 10 Hz and 500 Hz, with a gain of 500, and then sampled at 2048 Hz.

Before the experiment, the participant's forearm length is measured as the distance between the olecranon process and the ulnar styloid process. The wrist circumference is measured at

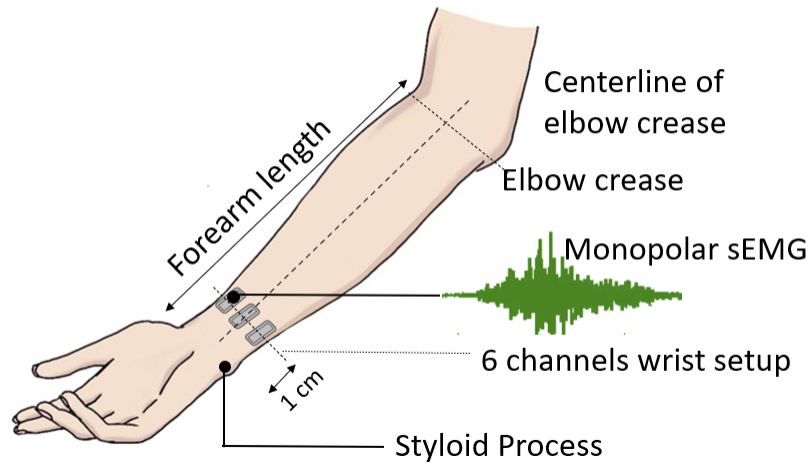


Figure 6.1: Positions of the six electrodes on the wrist (dorsal view, three electrodes are on the posterior wrist surface). The electrodes were equally spaced around the wrist at a distance of one cm from the styloid process. The monopolar EMG of each ring was acquired for subsequent processing and analysis.

one cm away from the ulnar styloid process in the proximal direction. After taking these measurements, the electrodes are placed on the wrist. Prior to electrode placement, the skin surface was shaved to remove hairs, cleaned with an alcohol swab, and abraded with a paper towel. Six monopolar EMG electrodes (AM-N00S/E, Ambu, Denmark) were placed in the form of a ring, equally spaced around the wrist. A detailed pictorial representation is provided in [Figure 6.1](#). To maintain consistency of the positions of the electrodes across all participants, the first electrode in the ring was anatomically positioned on the centerline of the elbow crease and the rest of the electrodes were numbered in clockwise order from one to six as shown in [Figure 6.1](#) [62, 105]. The electrodes were part of a multi-ring forearm and wrist setup as described in the data descriptor which also provides additional details on the acquisition setup [56, 107].

6.2.3 Experimental Protocol

After the completion of the experimental setup, the participant is seated comfortably on the chair with both their upper limbs in a resting position. Visual instructions for performing the gestures were provided on the computer screen placed in front of the participants. The following 16 hand and wrist gestures were included in the current study: Lateral prehension (LP), thumb adduction (TA), thumb and little finger opposition (TLFO), thumb and index finger opposition (TIFO), thumb and little finger extension (TLFE), thumb and index finger extension (TIFE),

index and middle finger extension (IMFE), little finger extension (LFE), index finger extension (IFE), thumb extension (TE), wrist flexion (WF), wrist extension (WE), forearm supination (FS), forearm pronation (FP), hand open (HO), and hand close (HC). A pictorial representation of the gestures is provided in Figure 3.2. The order of the 16 gestures was randomized and a resting (REST) trial was collected after all 16 gestures were performed once. A ten-second relaxing period was provided between each trial. One continuous data acquisition of 17 gestures, including the REST, is called one run. Each subject performed seven runs, resulting in 119 trials or contractions (17 x 7). The subject could also request additional rest when he/she felt necessary.

6.2.4 EMG Signal Processing

The monopolar EMG signals from the six channels were first re-referenced by a common average procedure where for every sample, the mean of the six channels was subtracted from each channel. The processed signals were then segmented into 200 ms width windows, with a 150 ms overlap. Each window was then processed using the frequency division technique (FDT) feature extraction [109]. This method calculates the magnitude of L frequency bands. For the i^{th} band, let $f_{i,1}$, and $f_{i,E}$ denote the frequency values of the two endpoints. As such, for each window, the i^{th} feature is calculated as:

$$FDT_i = F \left[\sum_{j=1}^{n_i} |X(f_{i,j})| \right], i = 1, 2, \dots, L, \quad (6.1)$$

where $X(\cdot)$ denotes the magnitude of the FFT spectrum, and $F[\cdot]$ denotes a logarithm operator to obtain a smooth value for better classification results. In the current study, the whole EMG frequency band (20-450 Hz) is subdivided into six equal-width frequency bands: 20-92, 92-163, 163-235, 235-307, 307-378, and 378-450 Hz, consistent with prior studies [47]. Therefore, for a single sample recorded from the six channels, the feature vector extracted from each window comprises 36 FDT features. For each trial recording of five seconds long, the extracted features result in $P \times D$ matrix where P equals 33 and D equals 36. Multiple trials from different days were used for training (termed as enrollment) of biometric models as discussed in a later section (Section II.H). The testing data involved biometric matching as described below.

The Mahalanobis distance, which takes into consideration the correlation between features, was considered for biometric comparisons [62, 105, 106]. In these studies, the Mahalanobis distance was robust to the high similarities between the EMG signals from neighboring channels. For a given feature vector sample p , which is the input from a specific user (the claimant) while

performing a specific gesture, its matching score, $S_{i,j}$, with the i^{th} gesture and the j^{th} user, was defined as the Mahalanobis distance between the sample and the class centroid:

$$S_{i,j}(p) = \sqrt{(p - \mu_{i,j}) \Sigma_{i,j}^{-1} (p - \mu_{i,j})^T} \quad (6.2)$$

where $\mu_{i,j}$ is the centroid of the gesture of the class and the user, and $\Sigma_{i,j}$ is the covariance matrix for the specific gesture and user class. Both the centroid and covariance were estimated from the enrollment data based on the within-day and cross-day analysis (details in Section II.H).

6.2.5 Multi-code Biometric Framework

A standard biometric system consists of five modules: 1) sensor module which collects the biometric data, 2) feature extractor for generating feature vectors utilized as biometric entries, 3) matcher module that compares with the genuine user's template to generate a score, 4) ranking module that produces a rank after sorting the scores for all the users and 5) decision module to grant access or rejection based on a pre-set threshold [68]. In the authentication mode, the ranking module is absent and hence the decision is made on the scores (access granted or rejected). In the identification mode, the usual practice is that the K identities with the top K highest scores are stored, and if the claimed identity is one of these K identities, authorization of the claim is rendered. Fusion strategies at different modules for EMG-based biometric authentication have been investigated, and the decision-level fusion approach was found to produce the best overall performance [106]. The performed hand/wrist gestures are treated as codes in the context of EMG biometrics. An example of a user's authentication code sequence of codelength M can be denoted as $[C_1, C_2 \dots C_m \dots, C_M]$, where C_m is the m^{th} code (gesture). For the analysis presented below, 50 random sequences were generated by combining randomly selected four gestures out of the 16 gestures performed by each participant. A decision-level fusion of each code sequence was employed as described below, and the evaluation metrics were compared for a single-code ($M=1$) and a multi-code configuration ($M=4$) based on previous findings [106]. For the identification mode, an additional parameter *rank-K* was analyzed by varying the value of K ($K=1-15$) for the single-code and multi-code configurations for each code sequence.

A weighted majority scheme was used for the M codes of the multi-code framework. In the authentication mode, the extracted feature from claimant data is compared to the corresponding template in the database. Hence, the authentication result is binary: 1 (for genuine user) and 0 (for impostor).

The identification mode is a multi-class problem with J comparisons between the claimant and all the templates in the database (total users = J). However, only when the claimant is within

the top K users, a positive identification will be rendered. For M codes, let $d_{m,j}$ be the degree of certainty or the m^{th} code (C_m) and j^{th} user defined as

$$d_{m,j} = \begin{cases} 1, & \text{if code } m \text{ is correct for } j_1, j_2, \dots, j_k, \\ 0, & \text{otherwise.} \end{cases} \quad (6.3)$$

The discriminant function for each user ‘ j ’ obtained through the weighted voting is

$$g_j = \sum_{m=1}^M w_m d_{m,j} \quad (6.4)$$

where w_m is the weight attached to the m^{th} code. Here w_m is the single-code recognition accuracy of each gesture, averaged over the three days. Based on the value of M , w_m is then normalized to 1.

$$\sum_{m=1}^M w_m = 1 \quad (6.5)$$

In the case of an authentication system, the claimant matches with the template if the discriminant g has a majority ($> 50\%$). For the identification system, the claimant is assigned with the template with the highest value of g . The weight w_m was previously determined by the recognition accuracy of individual gestures which are listed in [Appendix C](#).

6.2.6 Authentication and Identification Evaluation

The false acceptance rate (FAR) and the false rejection rate (FRR) were calculated to evaluate biometric authentication. FAR is the rate of accepting an impostor, and FRR is the rate of rejecting a genuine user. In principle, the FAR and FRR should be as small as possible for biometric applications. The detection error tradeoff (DET) curve is the relationship between the FAR and FRR . The equal error rate (EER) is the point on the DET curve where the FAR is equal to the FRR . The EER is a commonly used authentication metric that can be used to compare the performance of different biometric traits: The lower the EER value, the better the performance. Additionally, the area under the DET curve (AUC) value was also evaluated. For an accurate assessment of the biometric authentication capacity of the EMG biometrics, two common authentication scenarios were investigated: 1) Normal Test: where the correct code sequence was only known to the genuine user, while the impostor had no knowledge of the code sequence and presented

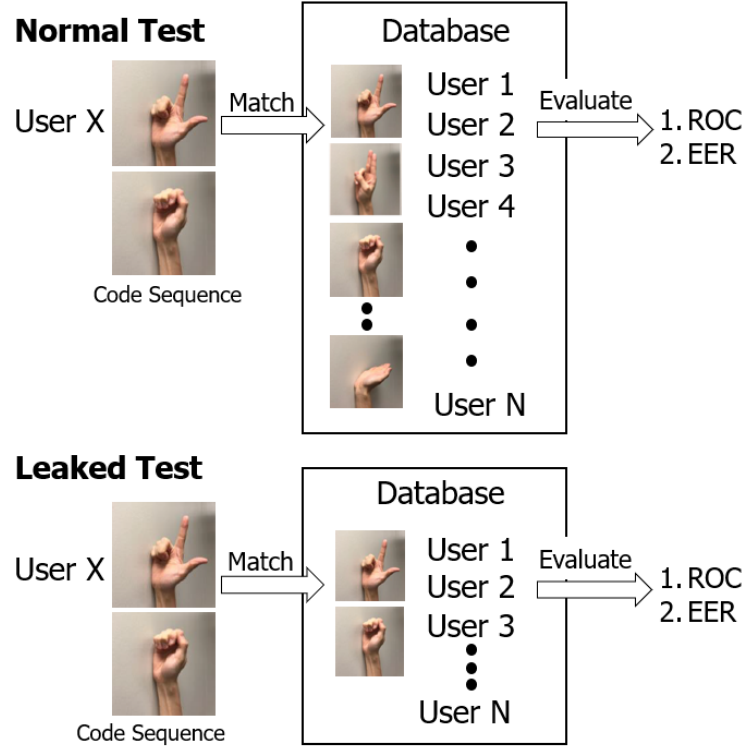


Figure 6.2: The normal test and the leaked test scenarios for biometric authentication. In both the scenarios the target user's identity is known to the claimant. The access code is unknown to the claimant in the normal test scenario, while it is known to the claimant in a leaked test scenario.

a random sequence different from the one used by the genuine user; 2) Leaked Test: where the correct code sequence for the genuine user was compromised, and the impostor presented the correct code sequence by performing the corresponding gestures. This is the scenario where the knowledge-based security is completely compromised, and the system is solely dependent on the biometric security.

In both cases, for the m^{th} code in a single-code ($M=1$) or multi-code ($M=4$) configuration, the genuine score, G_m , was obtained from the authentication gesture C_m of the genuine user. In the normal test, the impostor score I_m was obtained from the other gestures performed by other users. For the leaked-test scenario, I_m was obtained from C_m for all the other users. The weighted majority decision fusion scheme was implemented using G_m and I_m to obtain the final FAR_M , FRR_M , and EER_M .

For the identification mode, the *rank-K* accuracy is defined as the probability of correctly

predicting the genuine user within the top K likely users returned by the system for a code sequence of code length M . As the identification is a 1: J comparison (J = total users), the unknown presenting user is compared to J templates in the database, and J scores are generated. The returned identities are sorted according to the scores and the top- K users are selected. The value of *rank- K accuracy* is proportional with respect to the value of K , *i.e.*, the higher the K , the more likely it is for the genuine user to be returned. Their relation was summarized by the cumulative match characteristic (CMC) curve, which plotted rank- K against K , where K varied from 1-15. This analysis was repeated 50 times for randomly generated code sequences for the single-code and multi-code configurations as described previously in Section II.E. For both the authentication and identification mode, a cross-validation scheme corresponding to the multiple days and trials was incorporated as described in Section II.H. As the commonly used metrics, the values of *rank-1*, *rank-3* and *rank-5 accuracy* were reported in this study.

6.2.7 Intruder Analysis for Identification mode

The standard identification analysis does not include the scenario in which an intruder is an unregistered user. For such a scenario, an “authentication + identification” approach is taken, where the intruder can be rejected by a threshold-based authentication stage before the identification stage [60]. This threshold-based authentication stage introduces the threshold value ($Th_{i,j}$) for the i^{th} gesture and the j^{th} along with template parameters $\mu_{i,j}$, and $\Sigma_{i,j}$. From N training samples, the threshold value is defined as the upper limit of the scores obtained by matching with the template mean and covariance (Eqn 2)

$$Th_{i,j} = \max(\{S_{i,j,1}, S_{i,j,2}, \dots, S_{i,j,N}\}) \quad (6.6)$$

Decreasing the threshold value reduces the number of intruder matches, however, the correct identification of the true user also decreases. Based on preliminary results and due to its simplicity, the maximum training score was chosen as the threshold value. For the intruder analysis, score $S_{i,j}$ was obtained from Eqn. 2 by matching an unknown user with a template, which is further compared to the threshold $Th_{i,j}$

$$S_{i,j} = \begin{cases} S_{i,j}, & \text{if } S_{i,j} \leq Th_{i,j}, \\ \text{False}, & \text{if } S_{i,j} > Th_{i,j}. \end{cases} \quad (6.7)$$

For a registered user, only the true matches (Eqn. 3) were returned, and the scores were further sorted. The *rank-5 accuracy* was estimated similar to the standard identification analysis

as described previously. In the case of an intruder, successful rejection (Eqn. 3) from all the registered users' contributed to the *rank-5 accuracy*. A different evaluation metric, the intruder rejection rate (*IRR*) was defined as the ratio of correct rejections after matching an intruder's data with all the registered users

$$ORR = \frac{\left[\sum_{i=1}^N S_j == \text{False} \right]}{\sum j}, j = 1, 2, \dots, J \quad (6.8)$$

For the intruder analysis, the data from 42 users were registered to the biometric system and one user was considered an intruder. This was repeated 43 times in a leave-one-out cross-validation scheme and the *rank-5 accuracy* and *IRR* were reported. This was repeated 50 times for randomly generated code sequences used as biometric tokens for identification. Further, this entire analysis was repeated in cross-validation schemes corresponding to the multiple days and trials, which were incorporated as described in the following section. The purpose of the intruder analysis is to assess the performance degradation in the presence of unregistered users and the threshold-based rejection schemes for reducing such intrusions.

6.2.8 Multi-day Analysis

In the current study, data was collected from each user over three different days comprising seven trials each and 16 gestures in each trial. The biometric authentication and identification performance evaluation was investigated in three multi-day analyses: within-day analysis and two separate cross-day analyses, namely single cross-day (SCD) and cumulative cross-day (CCD). For the within-day analysis, six trials of the gestures each day were used as enrollment data (training), and the remaining one trial of that day was used as claimant data (testing), resulting in a leave-one-out (LOO) cross-validation scheme, equivalent to seven-fold cross-validation. The biometric performance for each fold was estimated as described in Sections II.F. The cross-validation was repeated for each of the three days and the average performance metrics were reported.

For the single cross-day analysis, six trials of the gestures from one day were used as the enrollment data and the data from one trial from each of the remaining two days were used as the claimant data. This step was repeated seven times by varying the enrollment and claimant trials from the specific days and thus resulting in between-day seven-fold cross-validation. The cross-validation was repeated three times each day and the average biometric performance was reported.

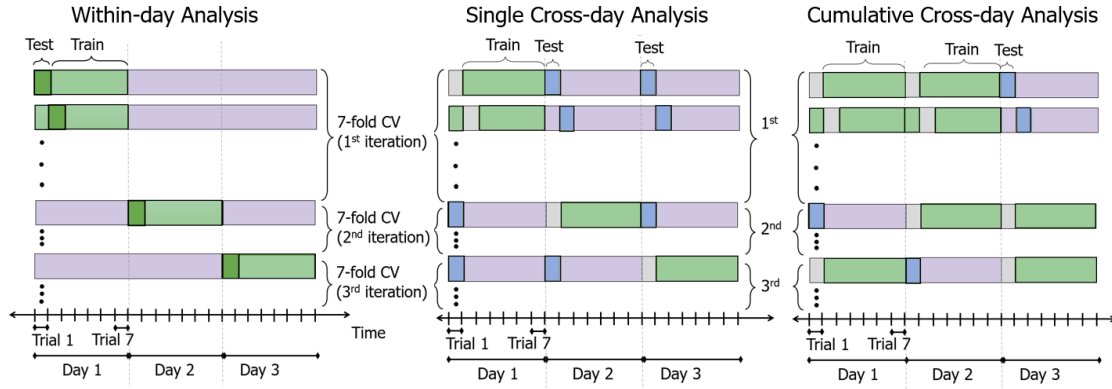


Figure 6.3: (From left to right) Within-day, Single cross-day and Cumulative cross-day analysis. The corresponding training (enrollment) data for each analysis are represented in green; the testing (claimant) data are represented in dark green (for within-day analysis) and blue (for cross-day analysis). The x-axis represents the timeline of the study consisting of multiple days; the x-axis tick marks represent the different trials performed in one day. Each of the analysis is repeated for the three days.

For the cumulative cross-day analysis, six trials of the gestures from two of the three days were used as the enrollment data and the data from one trial from the remaining one day was used as the claimant data. Seven-fold cross-validation for all seven trials was implemented by varying the enrollment and claimant trials from the specific days. The cross-validation was repeated three times each day and the average biometric performance was reported. A graphical representation of the within-day, single cross-day, and cumulative cross-day analysis is provided in [Figure 6.3](#).

6.2.9 Statistical Analysis

The study aimed to investigate the multi-day biometric performance of the wrist EMG biometric system. The performance of a multi-code EMG system was compared to the single-code configuration. For each of the three analysis scenarios, *i.e.*, within-day, single cross-day, and cumulative cross-day, a Wilcoxon rank sum test was performed on the *EER* and *AUC* of the authentication mode, *rank-1*, *rank-3* and *rank-5 accuracy* of the identification mode, *rank-5 accuracy* and *IRR* of the intruder analysis to determine if there was any significant difference between the single-code and multi-code configuration. To compare the performance of each metric between the three multi-day analyses (WD, SCD, and CCD), pairwise comparisons were performed using the Wilcoxon rank sum test, while keeping the configuration level (single-code and multi-code)

fixed. Additionally, for the intruder analysis, the Wilcoxon rank sum test was used to compare the *rank-5 accuracy* to the corresponding values for the standard identification analysis. All statistical tests were performed using RStudio 1.0. 136 (RStudio, Boston, MA).

6.3 Results

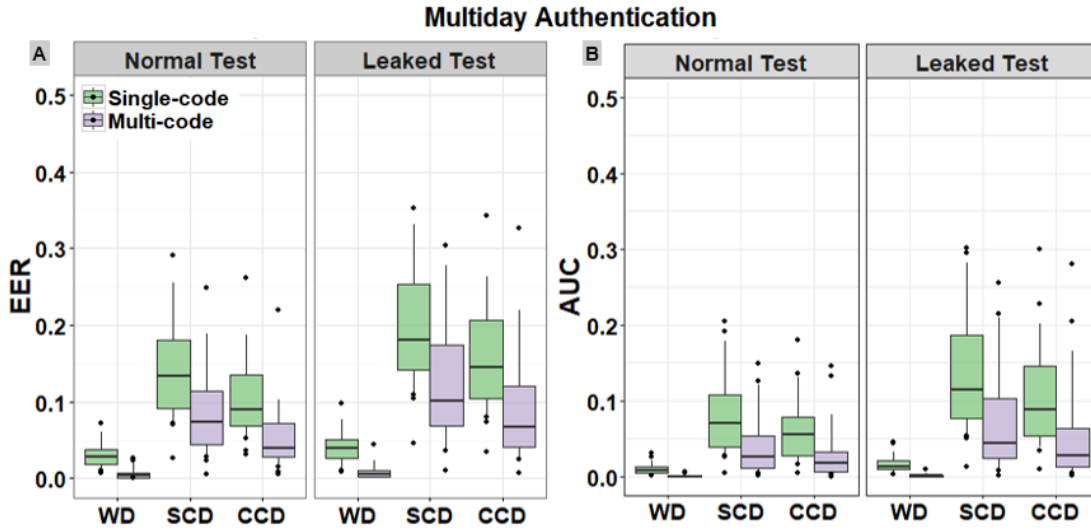


Figure 6.4: Biometric authentication and identification performance. For the multi-day authentication performance (Fig 4A), the single-code and multi-code *EER* values for the normal test and leaked test scenario are shown. The x-axis represents the three multi-day analyses: within-day (WD), single cross-day (SCD), and the cumulative cross-day (CCD) and the y-axis represents the performance error. Each boxplot represents the interquartile range (IQR, 25th – 75th percentile) and the center horizontal line represents the median *EER* value. The whiskers (solid vertical lines) represent the datapoints within the $1.5 \times \text{IQR}$ threshold. The outliers (solid black circles) are defined as those individuals whose *EER* was greater than the $1.5 \times \text{IQR}$ threshold. For the multi-day identification, the *rank-1* and *rank-5 accuracy* for the single-code and multi-code configuration are shown. The x-axis represents the three multi-day analyses, and the y-axis represents the rank-K accuracy where $K = 1, 5$.

Figure 6.4 shows the biometric performance in a single-code and multi-code configuration for the three timeline-specific analyses (termed as multi-day analysis for further representation):

within-day, single cross-day, and cumulative cross-day— for the biometric authentication. Figure 6.4 A shows the *EER* and Figure 6.4 A shows the *AUC* distribution in a normal test and leaked test scenario of the authentication mode. Figure 6.5 presents rank-1, rank-3 and rank-5 accuracy of the identification mode. The individual user performance for the identification mode (rank-5 accuracy) for the three multi-day analyses is shown in Figure 6.6. Figure 6.7 shows the DEC curves for the two scenarios of the authentication mode and the CMC curve for the identification mode.

As expected and evident from Figure 6.4, the within-day performance was significantly higher ($p < 0.001$) than the other two timeline-specific analyses. Further, the median *EER* of the cumulative cross-day analysis was significantly lower than the single cross-day analysis ($p < 0.001$) which is described in Section III.C.

While comparing CCD performance to SCD performance, an overall dominance of CCD was observed for the testing scenarios (normal test, leaked test, and identification). Specifically, for the multi-code configuration in a normal test scenario, the median *EER* (0.068, Q1=0.041, Q3=0.125) for the CCD was significantly lower ($p < 0.001$) than the corresponding median *EER* (0.039, Q1=0.029, Q3=0.072) for SCD analysis. For the identification mode, the multi-code configuration resulted in a median rank-5 accuracy of 93.0% (Q1=82.3%, Q3=98.0%) for the CCD which was significantly higher ($p < 0.001$) than the corresponding SCD median rank-5 accuracy of 85.1% (Q1=66.6%, Q3=91.6%). This result was particularly encouraging, as it indicates the training data from two different days were homogeneous enough to provide significant benefits for biometric authentication, providing a basis for further incremental adaptive training. For each of the three timeline-based analyses and the two biometric application modes, the detailed analysis of single-code and multi-code performance is presented below.

6.3.1 Within-day analysis

For the authentication mode, the within-day performance of the single-code ($M=1$) configuration in the normal and leaked test scenarios had a median *EER* of 0.028 (Q1=0.018, Q3=0.038) and 0.04 (Q1=0.026, Q3=0.052), respectively. The performance of the multi-code framework ($M=4$) was significantly improved ($p < 0.001$), as expected and the performance reached a median *EER* value of 0.004 (Q1=0.001, Q3=0.008) for the normal test and median *EER* of 0.006 (Q1=0.002, Q3=0.011) for the leaked test scenario. A similar trend was observed for the median *AUC* values (Figure 6.4 B). The median *AUC* values for the multi-code framework were 0.001 (Q1= 10^{-5} , Q3=0.002) in the normal test scenario and 10^{-5} (Q1= 10^{-5} , Q3=0.001) in the leaked test scenario.

For the identification mode, the within-day performance of the single-code configuration reported a median rank-1 accuracy of 90.0% (Q1=86.9% and Q3=91.4%) and median rank-5

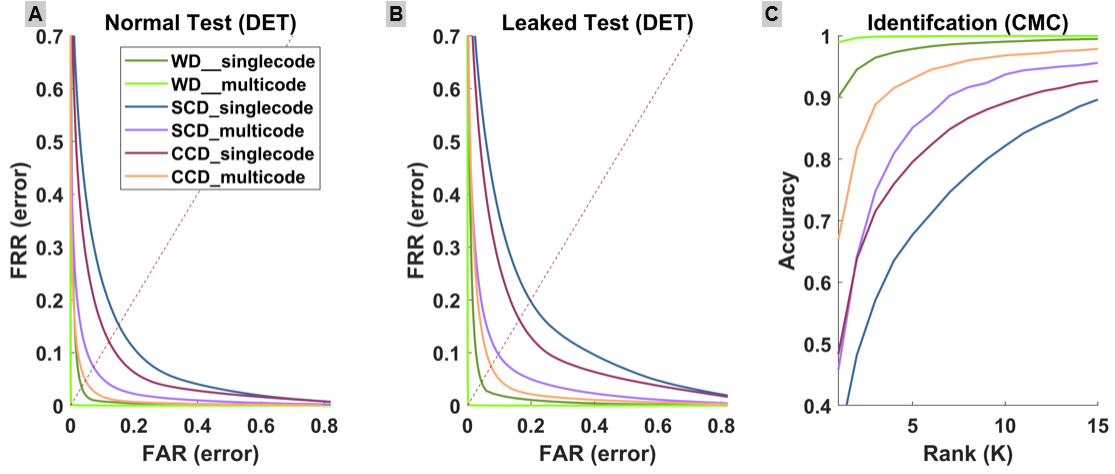


Figure 6.5: DET and CMC curves for the biometric tests. The averaged detection error DET curves for the two authentication scenarios: normal test (Fig 5A) and leaked test (Fig B), in three multi-day analyses: within-day (WD), single cross-day (SCD) and cumulative cross-day (CCD), and the two configurations: single-code and multi-code. The x-axis represents the FAR values, and the y-axis represents the FRR values. The averaged CMC curves (Fig 5B) are shown for the identification mode in the three multi-day analyses and the two configurations. The x-axis represents the Rank (K) of the identification system, and the y-axis represents the identification accuracy.

accuracy of 97.9% (Q1=97.3%, Q3=98.4%). For the multi-code configuration, the median *rank-1* accuracy was —99.0% (Q1=98.1% and Q3=99.5%) and the median *rank-5* accuracy was 99.9% (Q1=99.8%, Q3=99.98%) which was significantly higher than the corresponding single-code performances ($p < 0.001$).

6.3.2 Single Cross-day Analysis

For the authentication mode in an SCD analysis, the single-code configuration had a median *EER* of 0.133 (Q1= 0.091, Q3=0.191) for the normal test scenario, which was significantly higher ($p < 0.001$) than the median *EER* (0.72, Q1=0.044, Q3=0.115) of the multi-code configuration. Similarly, for the leaked test scenario, the single-code configuration resulted in a median *EER* of 0.181 (Q1=0.136, Q3=0.256) which was significantly higher ($p < 0.001$) than the median *EER* (0.101, Q1=0.068, Q3=0.189) of the multi-code configuration. A similar trend was observed for the median *AUC* values (Figure 6.4 B). The median *AUC* values for the multi-code framework

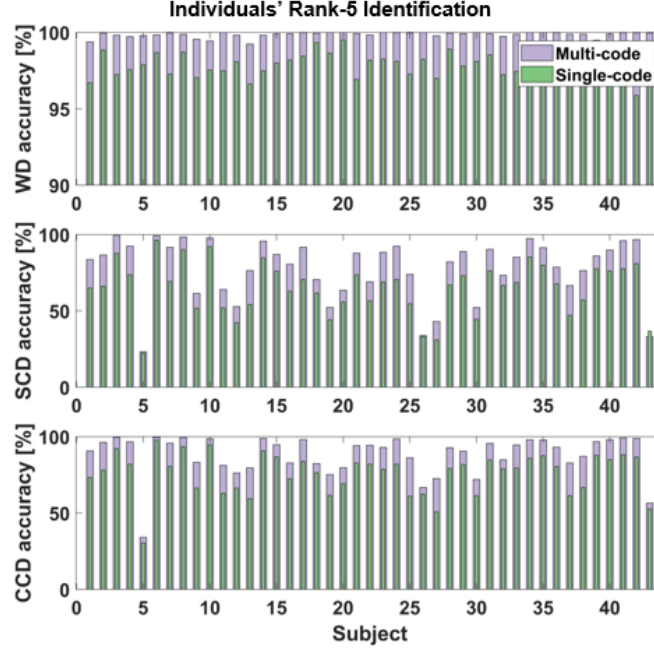


Figure 6.6: Individual identification performance. The *rank-5 accuracy* for all the users for the WD (top), SCD (middle) and CCD (bottom) analysis is illustrated. The multi-code values are shown in violet color and the corresponding single-code values are shown in green color.

were 0.026 (Q1=0.011, Q3=0.053) in the normal test scenario and 0.043 (Q1=0.024, Q3=0.103) in the leaked test scenario. As expected, the median *EER* values for the leaked test were significantly higher ($p < 0.001$) than the median *EER* values of the normal test, for both the single-code and multi-code configurations.

For the identification mode in an SCD analysis, it was found that significantly better performances ($p < 0.001$) were obtained for the multi-code configuration than for the single-code configuration. The multi-code configuration reported a median *rank-1 accuracy* of 45.7% (Q1=30.1%, Q3=67.6%) and a median *rank-5 accuracy* of 85.1% (Q1=66.6%, Q3=91.6%) which was higher than the corresponding single-code *rank-1 accuracy* of 33.6% (Q1=21.0%, Q3=46.6%) and *rank-5 accuracy* of 67.7%, (Q1=54.1%, Q3=76.3%).

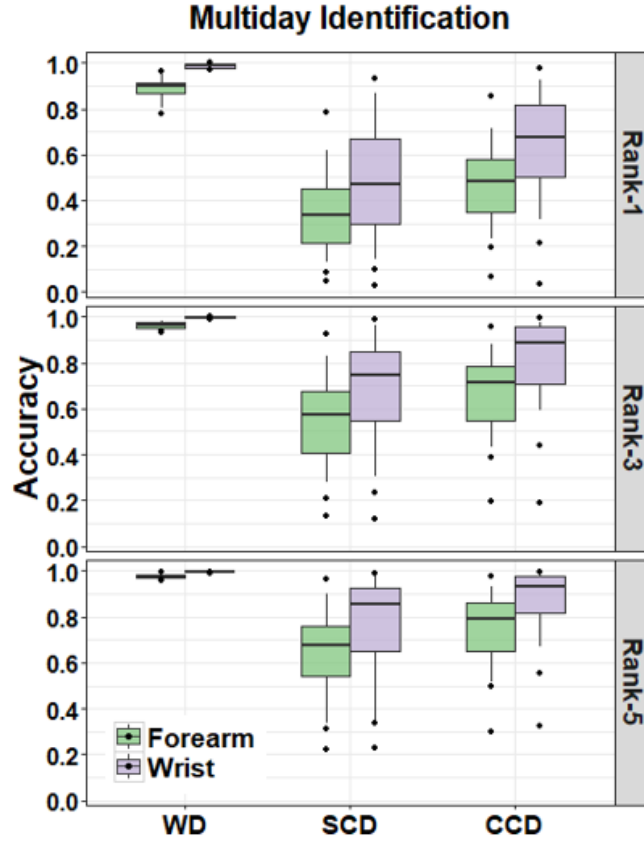


Figure 6.7: Biometric identification performance. The *rank-1*, *rank-3* and *rank-5* accuracy for the single-code and multi-code configuration are shown. The x-axis represents the three multi-day analyses, and the y-axis represents the rank-K accuracy where $K = 1, 5$. Each boxplot represents the interquartile range (IQR, 25th – 75th percentile) and the center horizontal line represents the median accuracy value. The whiskers (solid vertical lines) represent the datapoints within the $1.5 \times \text{IQR}$ threshold. The outliers (solid black circles) are defined as those individuals whose accuracy was greater than the $1.5 \times \text{IQR}$ threshold.

6.3.3 Cumulative Cross-day Analysis

For the authentication mode in the CCD analysis, the single-code configuration had a median *EER* of 0.09 ($Q1=0.069$, $Q3=0.136$) for the normal test scenario, which was significantly higher ($p < 0.001$) than the median *EER* of 0.039 ($Q1=0.029$, $Q3=0.072$) of the multi-code configuration. Similarly, for the leaked test scenario, the single-code configuration resulted in a median *EER*

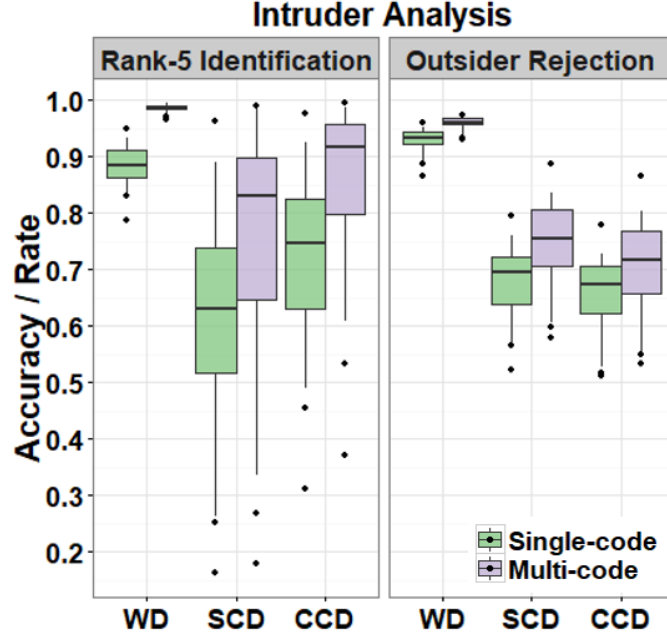


Figure 6.8: Intruder analysis for the identification mode. For the multi-day authentication performance, the single-code and multi-code *rank-5 accuracy* and *IRR* values are shown. The x-axis represents the three multi-day analyses: within-day (WD), single cross-day (SCD), and the cumulative cross-day (CCD) and the y-axis represents the *rank-5 accuracy* (left), and the *IRR* (right). Each boxplot represents the interquartile range (IQR, 25th – 75th percentile) and the center horizontal line represents the median *rank-5 accuracy/IRR* value.

of 0.145 (Q1=0.101, Q3=0.209) which was significantly higher ($p<0.001$) than the median *EER* (0.068, Q1=0.041, Q3=0.125) of the multi-code configuration. A similar trend was observed for the median *AUC* values (Figure 6.4 B). The median *AUC* values for the multi-code framework were 0.028 (Q1=0.012, Q3=0.063) in the normal test scenario and 0.017 (Q1=0.006, Q3=0.033) in the leaked test scenario. Similar to the WD and SCD analysis, the median *EER* values for the leaked test were significantly higher ($p<0.001$) than the median *EER* values of the normal test, for both the single-code and multi-code configurations.

For the identification mode in a CCD analysis, it was found that the multi-code configuration performed significantly better ($p<0.001$) than the single-code configuration. The multi-code configuration reported a median *rank-1 accuracy* of 66.9% (Q1=51.4%, Q3=80.9%), median *rank-3 accuracy* of 88.8% (Q1=70.9%, Q3=95.9%) and a median *rank-5 accuracy* of 93.0% (Q1=82.3%, Q3=98.0%) which was higher than the corresponding single-code *rank-1 accuracy*

of 48.4% (Q1=33.3%, Q3=58.1%), median *rank-3 accuracy* of 71.5% (Q1=54.8%, Q3=78.5%) and *rank-5 accuracy* of 79.5%, (Q1=66.2%, Q3=86.0%).

6.3.4 Intruder Analysis for Identification Mode

The effect of an intruder was analyzed for the identification mode in each of the three multi-day analyses. Figure 6.7 shows the *rank-5 accuracy* and the *IRR* in a single-code and multi-code configuration identification mode. As expected, it was observed that the *rank-5 accuracy* decreased in all three multi-day analyses (WD, SCD, and CCD), compared to the standard identification analysis. Specifically for the CCD scenario, the multi-code configuration reported a significant decrease ($p<0.001$) in median *rank-5 accuracy* from 93.0% (Q1=82.3%, Q3=98.0%) for the standard analysis to 91.7% (Q1=79.3%, Q3=96.5%) for the intruder analysis. For the SCD scenario, the multi-code configuration reported a significant decrease ($p<0.001$) in median *rank-5 accuracy* from 85.1% (Q1=66.6%, Q3=91.6%) for the standard analysis to 82.9% (Q1=61.7%, Q3=90.6%) for the intruder analysis. For all three multi-day analyses, the multi-code configuration had significantly higher *rank-5 accuracy* ($p<0.001$) than the single-code configuration.

For the multi-code configuration, the median *IRR* was significantly higher ($p<0.001$) for the WD (96.1%, Q1=95.6%, Q3=96.8%) than the SCD (75.5%, Q1=70.4%, Q3=80.8%) which was significantly higher than the CCD (71.7%, Q1=65.7%, Q3=77.1%) analysis. For the CCD analysis, it was observed that the multi-code configuration had a significantly higher ($p<0.001$) median *IRR* (71.7%, Q1=65.7%, Q3=77.1%) than the single-code configuration (67.3%, Q1=62.1%, Q3=70.7%). This was also consistent with SCD and WD analysis. For all three multi-day analyses, the multi-code configuration had significantly higher *rank-5 accuracy* ($p<0.001$) than the single-code configuration.

6.4 Discussion

The study presented the multi-day performance of biometric authentication and identification on the largest EMG hand-gesture dataset. The single-code configuration was compared to the multi-code configuration as the latter facilitates the fusion of hand gestures for improved biometric performance. Two evaluation scenarios for the authentication mode, i.e., the normal and leaked test, and the identification mode were investigated. The normal test scenario corresponds to the dual-security comprising both biometric-level (i.e. individual-specific characteristics in EMG signal) and knowledge-level (specific hand-gesture/code sequence). The leaked test scenario corresponds to the situation when the knowledge-level security (code sequence) is compromised, and only the biometric-level security is retained. The identification mode corresponds to

the scenario where the system performs individual recognition from a database. Additionally, a more practical identification analysis involving intruders and their rejection was investigated. A within-day (WD) and cross-day, (SCD and CCD) analyses were performed for each of the comparisons and the results demonstrated the biometric performance, as discussed in the following sections.

6.4.1 Single-code vs multi-code biometric performance

The single-code (or the single gesture) configuration was compared to the multi-code (code sequence of four randomly selected gestures) configuration. As expected, there was an overall reduction in authentication *EER* and an increase in identification accuracy with the multi-code configuration than with the single-code configuration. These findings were in agreement with multiple password-based studies using EMG from gestures [60–62, 64]. The findings in the current study were also consistent with a previous study that performed only the within-day analysis on a forearm setup [106]. For all the codelengths, it was observed that the leaked test scenario had a higher *EER* than the normal test scenario, thus suggesting that an improved performance is achieved by leveraging the knowledge-level security mode of EMG, *i.e.* with user-defined code sequences. This is a unique advantage of EMG-Biometrics as compared to other biosignals such as EEG and ECG, for which such knowledge-level security features is not available [106]. While the identification mode inherently does not allow customized codes for individuals, it facilitates the fusion of multiple codes to form a universal code sequence which can improve personal recognition.

6.4.2 Multi-day analysis

For both the single-code and multi-code configurations, it was observed that there was significant degradation in the authentication and identification performance in the cross-day analysis as compared to a within-day analysis. This is expected as the EMG signals are affected by non-stationary factors such as electrode shift, skin conditions (dry vs sweat), and physical conditions (rested vs exercising) [118]. While comparing the two cross-day analyses for the authentication mode, the CCD authentication *EER* was significantly lower than the SCD authentication *EER* for both the normal and leaked test scenarios. Similarly, it was observed that CCD identification accuracy was significantly higher than SCD identification accuracy. This suggests that EMG data from different days do have enough homogenous information such that training with data from multiple days improves the biometric authentication performance. For the CCD analysis (training data from multiple days), a multi-code configuration resulted in a median authentication *EER* of

Table 6.1: Biometric Traits Characteristics

Biometric	Spoofing	Customizable	Environment factors	Accuracy
Fingerprint	Easy	No	Robust	High
Iris	Easy	No	Sensitive	High
Face	Easy	No	Sensitive	High
Voice	Easy	Yes	Sensitive	Low
Gait	Hard	Yes	Sensitive	Low
Keystroke Dynamics	Hard	Yes	Robust	Low
Biosignals				
EEG	Hard	No	Sensitive	Low
ECG	Hard	No	Robust	Low
EMG	Hard	Yes	Sensitive	Acceptable

0.039 for the normal test and a median accuracy of 93% for the identification mode, which are comparable to most conventional biometrics as discussed in the following section. It was observed that some participants had an abnormally low identification performance, as compared to the rest of the participant pool (Figure 3.6). This could be due to the shift in electrode positions causing a mismatch in the feature vectors. Some strategies to address electrode shifts in future studies are discussed in Section IV.E. This would also help achieve higher *rank-1 accuracy* than the present study (66.9%). The cross-day analysis is crucial for evaluating a biometric system, as the results need to be consistent over multiple days. In the following sections, the CCD analysis (unless specified otherwise) will be used to discuss the biometric performance.

6.4.3 Comparison with state-of-the-art EMG Biometrics

A systematic comparison with EMG baseline studies has been challenging due to the differences in experimental protocol, feature extraction strategies, biometric analysis, and fusion strategies used. The WD performance was comparable to an identical forearm electrode setup [105, 106], suggesting that the wrist, a more convenient position for wearable devices can be considered for biometric applications. This finding was in agreement with another study where the wrist setup performed comparably if not better than the forearm setup in a gesture recognition application [8]. For the cross-day comparisons, we have focused on comparing the results of those biometric studies that incorporated a multi-day factor.

Biometric Authentication One study investigated a password-based EMG biometrics of codelength eight and reported a normal-test *EER* of 0.012 and the leaked-test *EER* of 0.1496 [64]. Another study involving a similar multi-code password of codelength 12 reported a normal-test *EER* of 0.0013 and the leaked-test *EER* of 0.0273 [61]. Another study incorporated a multi-day investigation, however, different codelengths were not investigated. They reported an average *EER* of 0.1634 for eight individual gestures [117]. While comparing these studies, the current study reported a median *EER* of 0.039 for the normal test and a median *EER* of 0.068 for the leaked test scenario which is in agreement. These findings can help further ongoing research on unlocking smartphones or developing wristband devices for biometric authentication [28, 126].

Biometric Identification Another identification study has investigated multiple protocols, codelengths, and electrode layouts for improving biometric performance [60]. For a fixed gesture sequence protocol and a codelength four (closest configuration to the present study), a mean *rank-1 accuracy* of 86.1% was reported which was higher than the current study (66.9%). This could be mostly due to: 1) the number of users in the previous study was 22 compared to 43 in the present study, and 2) the number of electrode channels was 256 compared to six in the present study. Another study with a similar setup of 22 users and 256 channels reported a *rank-1 accuracy* of 85.8%, however, the gesture task of each participant was different, and they reported a decrease in accuracy (<50%) with more than 30 inactive electrodes [62]. This further shows that the increased *rank-1 accuracy* could be due to the greater number of channels. The high-density electrode setup (256 channels) can have multiple limitations such as practical viability, inactive channels, and model training time, which are comparatively low while using a 6-channel setup. Another study involving a similar setup (22 gestures and 256 channels) and a codelength of eight used a deep learning model and reported a *rank-1 accuracy* of 87.2% and a *rank-5 accuracy* of 91.2% [30]. The present study reported a slightly higher *rank-5 accuracy* of 93%. The performance was also comparable to two other studies that involved multi-day testing data from 5 subjects and reported an identification accuracy of 74.1% and 93% [75, 117].

Protocols with intruder analysis To the best of our knowledge, only one study has investigated the effect of intruders on identification performance [60]. They have reported a decrease in identification performance from 99% to 93.1% while introducing unregistered users to the system. The experimental factors included 22 users, a codelength of 12, and 256 channels. The present study reported a decrease in *rank-5* identification accuracy from 93% to 91.7% for a codelength of four, which was in agreement with the previous study. Additionally, we observed that for a database of 42 users (one user removed as an intruder), an intruder is more likely to find a match with at least one of the registered users. We introduced a metric *IRR* to quantify the number of threshold-based rejections for each multi-day scenario. It was reported that for a

multi-day analysis, an $IRR > 70\%$ was achieved. The results suggest that the intruder encountered rejection from most of the registered users' threshold values. Reducing the threshold limit might improve intruder rejection, however, the genuine user might not be correctly identified. Additional research could explore approaches to enhance both identification performance and intruder rejection, ultimately leading to a more practical and robust biometric system.

6.4.4 Comparison with other biometrics

Conventional biometric traits such as fingerprint and facial recognition have already been widely implemented in daily consumer applications. It can be difficult to compare with conventional benchmarks because research in this area has been ongoing for several decades [25]. Datasets including thousands of individuals are available online, which indicates the abundance of resources for researching such traits [12]. However, with advancing technology, data leakage and spoofing have become increasingly easier, affecting an important aspect of our digital society. Therefore, unconventional biometric traits such as biosignals, gait, keystroke dynamics, etc., have been investigated to address the limitations associated with conventional biometric traits [25]. The previous studies suggest that the biometric authentication performance of these traits ranges from $EER = 10^{-4}$ - 0.20 [25]. From this review, it was suggestive that an accurate biometric authentication $EER < 0.05$, while most of the novel biometric traits reported *rank-5 accuracies* $> 90\%$ [25]. The findings of the current study (authentication $EER = 0.039$ and identification *rank-5 accuracy* = 93%) suggest that wrist EMG-Biometrics resulted in accurate authentication and identification using a multi-code framework for combining hand gestures. This property of customizing a code based on knowledge can be achieved using some form such as voice, gait, and keystroke dynamics for improved security. This is an advantage over other biosignals such as ECG and EEG, which are highly difficult to customizable, if possible at all [105]. Overall, the biosignals are more difficult to spoof and are one of the main indicators of liveness detection [62], a security feature requiring the user was physically present while an authentication claim is attempted. While the EMG signals are affected by multiple non-stationary factors, the cross-day analysis suggests there are sufficient commonalities in multi-day EMG data for biometric authentication. A detailed comparison of EMG-Biometrics with the other common biometrics is provided in Table 6.1 [25]. As EMG-Biometrics holds distinct advantages over other common biometrics, further research involving wearable devices and extensive datasets will aid its social acceptance. Future research could record collect EMG as well as another conventional biometric (fingerprint/facial recognition) from individuals to facilitate easier comparison with baselines in the biometric field.

6.4.5 Limitations and future research directions

The results of the study addressed two major challenges in the current EMG-based research by analyzing a large sample size of participants and multi-day ($N=3$) recordings of EMG signals. However, there exist some limitations in the present study. Multiple force levels for the same gesture could not be collected. Due to the large sample size, it was not timely feasible to collect different levels of force. The participant was instructed to perform the gestures at a convenient force level (as they would normally perform a gesture in their daily lives). Furthermore, no marks were left on the participants' skin during the multiple sessions. While the electrode positioning was kept consistent to the best of our ability, there might have been some electrode shifts between days. Although a limitation, such a situation is more realistic than tightly controlling electrode positions. It can facilitate research into transfer learning approaches that are insensitive to electrode shift, and hence, beneficial for accurate biometric authentication [112]. A different study investigated EEG biometric characteristics such as permanence and uniqueness by analyzing six recording sessions over spanning over 3 years [88]. The inclusion of transfer learning in EMG biometric analysis can facilitate such a longitudinal investigation.

The 6-channel electrode setup was part of a bigger dataset, *i.e.*, GRABMyo, which includes 28 channels recorded from the forearm and wrist region [56, 107]. While preliminary results suggested no difference between the forearm and the wrist setup, further investigations could lead to novel findings [104]. Other strategies to improve the authentication and identification performance can be further investigated by implementing different fusion strategies [106]. Recently, deep learning studies have improved biometric performance and can be incorporated in similar analyses for more accurate results [9, 30, 49, 101].

6.5 Chapter Conclusion

The study presented the multi-day EMG Biometrics performance of the largest EMG-based hand gestures dataset (43 users $\times$ 3 sessions = 129 recordings) over the span of 30 days. A detailed multi-day analysis considering within-day and cross-day scenarios was performed for the feasibility of EMG-Biometrics as a novel trait for biometric authentications/identification. A multi-code biometric framework had superior performance than the single-code system, suggesting enhanced security with customizable gesture passwords. The authentication mode resulted in an *EER* value 0.039 for the normal test, which is comparable to conventional biometrics. The identification mode presented a *rank-5 accuracy* of 93% which is comparable to most novel biometric traits. In the presence of intruders, a threshold-based identification was able to achieve a *rank-5 accuracy* of 91.7%. The findings are close-to-practical due to the more realistic analyses

such as cross-day, leaked test, and intruder analysis. This could further facilitate the development of wearable EMG-based bracelets and wristbands for biometric authentication and identification purposes. The high cross-day performance was consistent over multiple days, a necessary characteristic for biometric authentication. Therefore, the presented findings could facilitate further research on EMG-Biometrics.

Chapter 7

MyoBM-Net: Two-stage Convolutional Feature Engineering for Long-term Biometric Authentication using Wrist Electromyography

This chapter contains unpublished materials. The chapter is currently under review in the *IEEE Journal of Biomedical and Health Informatics* as A. Pradhan A., He J., Tung J., and Jiang N., "MyoBM-Net: Two-stage Convolutional Feature Engineering for Long-term Biometric Authentication using Wrist Electromyography". In this chapter, a convolutional neural network and a novel training paradigm is proposed for accurate biometric authentication.

7.1 Chapter Introduction

Recently, surface EMG recorded from the forearm and wrist has been proposed as a novel biometric to address the limitations of conventional biometrics. Commonly used in prosthetic research, generated EMG signals encode unique patterns not only for different hand gestures but also for different users. This allows users to combine hand gestures as passwords with two-factor security: gesture- and biometrics-level components, which is an advantage over other biometrics [34, 64, 106]. One of the crucial features of a viable biometric trait is its longitudinal robustness to multi-day applications, where cross-session discrepancies affect the biometric performance. To address the performance degradation of EMG-based biometric methods in multi-day scenarios, robust EMG feature extraction methods are warranted.

The advancements in artificial intelligence have led to the adoption of deep feature extraction and classification methods in the biosignal domain. However, multiple studies use traditional feature extraction methods, which are not custom-trainable to a particular biometric application. With the advent of deep learning and more powerful computational resources, it is possible to utilize models such as convolutional neural networks (CNNs) for higher classification accuracy. Some studies have utilized convolutional feature engineering for both authentication and identification applications [3, 30, 49, 74, 127]. However, these feature engineering studies not only lacked a comparison to the traditional feature engineering methods but also the investigation of different architecture types for utilizing spatial information from EMG channels. Similar to two adjacent pixels in an image, two EMG electrodes placed close to each other share ample similar information. Convolutional filters can accommodate neighboring channels/pixels and utilize spatial information to the benefit of the classification task. In this study, we focus on the authentication mode, a more frequently used biometric application in our daily lives. We propose a convolutional feature extractor that utilizes a novel two-stage training paradigm. In the first stage, the network learns multi-user and multi-gesture features. In the second stage, the network is fine-tuned to specific user’s authentication application.

In the present study, we propose MyoBM-Net, a four-layered convolutional network that utilizes raw EMG signals for accurate biometric authentication. A novel two-stage training is proposed: 1) stage-1 training, where a generalized network is trained for multi-gesture and multi-user recognition, and 2) stage-2 training, where the network is fine-tuned for a user-specific authentication application. Therefore, the purpose of the study is two-fold: 1) to compare the authentication performance of convolutional and traditional feature extraction methods, and 2) to investigate the effect of different training paradigms and network architecture on authentication performance.

7.2 Methods

7.2.1 EMG dataset

The study used the GRABMyo, a large-sample multi-day EMG dataset, the details of which can be found in [107]. The dataset included EMG data from 43 healthy participants (23 M, 20 F), recorded from 6 monopolar electrodes (AM-N00S/E, Ambu, Denmark) placed around the wrist as shown in Figure 7.1. Data collection sessions were repeated on 3 days (Day 1, 8, and 29), where participants followed a similar protocol. In each session, participants performed seven repetitions of 16 static hand/wrist gestures followed by a REST trial. Therefore, the dataset consisted of 43 users, 3 days (spanning over a month), 16 gestures, and 7 repetition trials. The

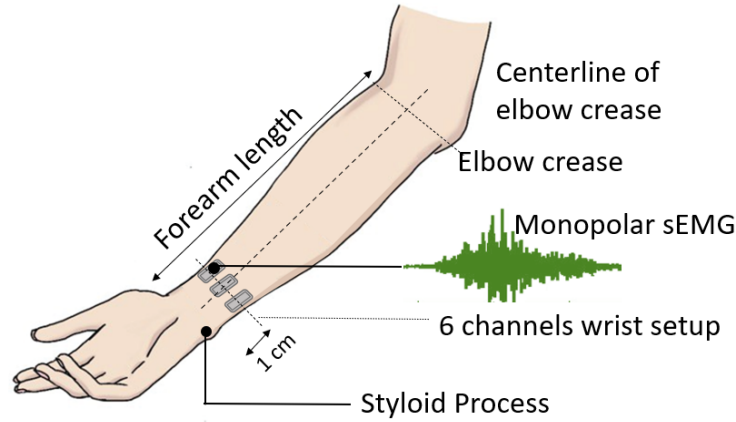


Figure 7.1: Positions of the six electrodes on the wrist (dorsal view, three electrodes on the posterior wrist surface are not shown). The monopolar EMG of each ring was acquired for subsequent processing and analysis. The diagram is obtained from the data descriptor [107].

comprehensive dataset facilitates a biometric analysis for assessing the uniqueness and long-term retainability of EMG-Biometrics. Additionally, a previous study by our group utilized the FDT feature extraction method on the GRABMyo dataset [108], providing the baseline performance for the proposed convolution-based feature extraction methods.

Cross-day training and testing dataset

The gesture performed by the users is termed as a ‘code’ in this study, and a combination of such codes is described as ‘a code-sequence’, or ‘a password’ in the context of biometrics. A code sequence of length four is chosen ($C=4$) for the present study. It is a sequence of four hand gestures. From our previous studies, two cross-day scenarios, namely single cross-day (SCD) and cumulative cross-day (CCD) scenarios were analyzed [103, 108] with slight modifications. The SCD scenario includes training data from either Day 1 or Day 2. The CCD scenario includes combined training data from both Day 1 and Day 2 and testing data from Day 3. For consistency in analysis, the testing analyses for both SCD and CCD used data from Day 3 only, which had a longer separation (at least 21 days apart) from the two training days. For the training data from each day, five out of seven trials were randomly selected. For testing, the remaining two trials were randomly selected. The analyses are repeated for thirty randomly generated code sequences ($N = 30$) as explained in further sections.

Signal pre-processing

Signal pre-processing was identical for both the training and testing phases. The monopolar sEMG signals from the six channels were re-referenced by common averaging. The processed signals were segmented into 200-ms windows with a step length of 150-ms (with 50-ms overlap). Each segment was then processed using the multiple feature extraction techniques discussed in sections II.B and II.C. In the training phase, the extracted feature set was the enrollment data for the genuine user. For the testing phase, the extracted feature set was the claimant data, which was used to compare with the enrollment data. A similarity score was generated from each comparison, or code-matching, and a score-based multi-code framework was used to generate a total score for the entire 4-code sequence [106]. The total score was compared to a threshold value, and a final authorization/rejection decision was rendered. Section II.E describes the matching method, testing scenario, and performance evaluation in further detail.

7.2.2 Traditional feature extraction method

In this study, the frequency division technique (FDT), a traditional frequency domain-based feature extraction method, was used for analysis [47, 109]. While primarily developed for prosthetic control, FDT has been utilized for EMG-Biometrics [105]. For FDT, the signals from each channel were divided into specific frequency sub-bands as given by:

$$FDT_i = F \left[\sum_{j=1}^{n_i} |X(f_{i,j})| \right], i = 1, 2, \dots, L, \quad (7.1)$$

where $X(\cdot)$ denotes the magnitude of the FFT spectrum, $F(\cdot)$ denotes a non-linear transformation (commonly a logarithmic operator) and L is the number of EMG channels. The signals from each channel were divided into specific frequency sub-bands of 20-92, 92-163, 163-235, 235-307, 307-378, and 378-450 Hz [109]. Fast-Fourier transformation (FFT) was performed in each sub-band, and the total power in each sub-band was estimated. The feature extraction was performed separately for each EMG channel. Therefore, the number of extracted features from six EMG channels was 36.

7.2.3 CNN-based feature extraction methods

CNN is a powerful deep-learning method in many EMG-based applications such as hand gesture recognition [45]. These networks perform temporal convolution operations of EMG signals and

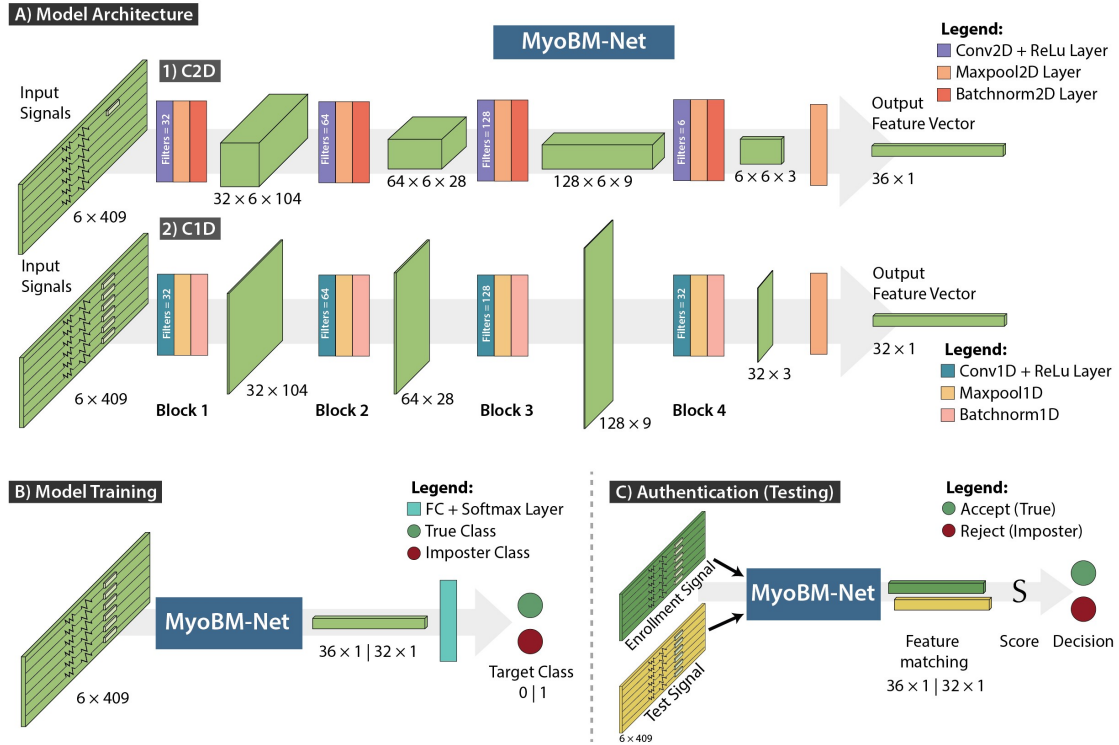


Figure 7.2: MyoBM-Net Convolutional Feature Extractor. Two network architectures (top-bottom) are proposed. The conv2d-based (C2D) utilized channel-specific information, while conv1d-based (C1D) used spatial information. Each architecture includes four blocks of convolution-pooling, normalization layers. The input consists of segmented EMG with 6 channels and 409 timepoints (corresponding to 250ms). The time-axis converges to a value of one (linear), which enables the extraction of relevant features. The output of the C2D and C1D architectures is a linear feature vector of length 36 and 32, respectively.

extract relevant features for a classification task. In a multi-channel EMG context, it is important to utilize the spatial information distributed across multiple EMG channels. In this study, MyoBM-Net, a convolutional feature extraction method, is specifically proposed for biometric authentication applications using EMG signals. Two variations of network architectures are investigated for MyoBM-Net: 1) 2D convolution layers performing within-channel operations (C2D) and 2) 1D convolution layers performing spatial operations (C1D). The two network architectures (Figure 7.2 A), network training (Figure 7.2 B) and the authentication application (Figure 7.2 C) are shown in Figure 7.2.

After the feature extraction step, the MyoBM-Net ('extractor' module) is coupled with a fully connected neural network classifier ('classifier' module) for network training (Figure 7.2 B). The extractor module involves four identical blocks of convolution layer (Conv2D/1D) with rectified linear (ReLU) activation, maximum value pooling layers (Maxpool2D/1D) followed by a batch normalization layer (BatchNorm2D/1D) (Figure 7.2 A). Hyperparameter tuning was performed using the grid search approach where different values for each hyperparameter were chosen iteratively and the best-performing model was obtained (shown in Table 7.1) [3]. The position of the batch normalization layers was determined by the hyper-parameter tuning. The BatchNorm2d layers at the end of each block provide regularization and stabilize the non-linear activation of ReLU layers, thus accelerating the training convergence. The pooling layers converge the time axis to a value of one, thus extracting only the most relevant features from each time window. At the end of the four blocks, the output of the extractor module is linearized to form feature vectors, which serve as the input to a fully connected classifier network with a SoftMax activation. The classifier module facilitates the training of the extractor module. The two variations (C2D and C1D) are discussed in the next subsections.

1) C2D The C2D utilizes conv2d layers, however, the kernel operation in each block is one-dimensional. Hence, the operations are performed within each channel. The kernel size on the time axis was set to seven (size = 1x7) in all the layers with a stride length of two. The number of convolution filters for the four blocks was set to 32, 64, 128, and 6, sequentially. Similar to the convolution layers, the maxpool kernels (size = 2) are limited to one dimension only. The output of the final pooling layer is a weight matrix of 6×6 (#EMG channels $\times$ #conv_4 filters). The matrix is linearized to obtain a final feature vector of length 36, which is the classifier's input.

2) C1D The C1D architecture uses conv1D layers, best suited for one-dimensional time series data. The conv1d requires a shape of C x L (C = Channel, L = Length) for the input data, while the width axis equals one (W = 1). The number of convolution filters for the four blocks was set to 32, 64, 128, and 32, sequentially. The conv1d layers perform operations in the time axis, however, the kernel size by default accommodates the input channels (C) of each convolution layer. The filter dimensions for the four convolution layers were 6 x 7, 32 x 7, 64 x 7, and 128 x 7. This architecture utilizes the spatial information of EMG channels. The final pooling layer results in a feature vector of length 32, which is input to the classifier.

The classifier module for the two architectures helps converge the training loss and thus facilitates network training. Predominantly, biometric comparisons employ distance-based matching, and hence, only the MyoBM-Net (extractor module) will be utilized for feature extraction and subsequent distance-based matching will be performed for biometric analysis.

Table 7.1: Network Hyperparameters

Hyper parameters	C2D	C1D	Grid-search values
Model Parameters			
Conv_1	Filters = 32	Filters = 32	{32, 64, 128, 256}
	Stride = (1, 2)	Stride = 2	{2, 3, 4}
	Kernel = (1, 7)	Kernel = 7	{3, 5, 7, 9}
	Pad. = (0, 7)	Pad. = 7	-
MxPool_1	Kernel = (1, 2)	Kernel = 2	{2, 3}
Conv_2	Filters = 32	Filters = 64	{32, 64, 128, 256}
	Stride = (1, 2)	Stride = 2	{2, 3, 4}
	Kernel = (1, 7)	Kernel = 7	{3, 5, 7, 9}
	Pad. = (0, 7)	Pad. = 7	-
MxPool 2	Kernel = (1, 2)	Kernel = 2	{2, 3}
Conv_3	Filters = 32	Filters = 128	{32, 64, 128, 256}
	Stride = (1, 2)	Stride = 2	{2, 3, 4}
	Kernel = (1, 7)	Kernel = 7	{3, 5, 7, 9}
	Pad. = (0, 7)	Pad. = 7	-
MxPool 3	Kernel = (1, 2)	Kernel = 2	{2, 3}
Conv_4	Filters = 6	Filters = 32	{32, 64, 128, 256}
	Stride = (1, 2)	Stride = 2	{2, 3, 4}
	Kernel = (1, 7)	Kernel = 7	{3, 5, 7, 9}
MxPool _ 4	Kernel = (1, 2)	Kernel = 2	{2, 3}
MxPool 5	Kernel = (1, 3)	Kernel = 3	-
Training Parameters			
Batch size	132	132	{66, 132, 264}
#eps S1	10	10	{5, 10, 50}
#eps S2	30	30	{66, 132, 264}
Optimizer	Adam	Adam	-
LR	0.001	0.001	$5e - 2 \rightarrow 5e - 4$

Conv: Convolution layer, MxPool: Max pooling layer, eps: Epochs, LR: Learning Rate,
C2D: network with conv2D layers, C1D: network with conv1D layers

user-specific authentication feature extractor (termed as S2). [Figure 7.3 A](#) explains the training paradigms in detail and [Figure 7.3 B](#) demonstrates the deployment of a trained network for the authentication analysis. The training method for the two stages is explained in the following subsections.

S1 Training

The objective of the S1 training is to extract features from EMG signals that contain both user-specific as well as gesture-specific information. For the training of such a network, we utilize all the users ($n = 43$) and the gestures ($n = 16$) from the training day (Day 1/2 for SCD and Day 1,2 for CCD). An episode-based training is employed where each episode is defined as randomly selecting 35 users (out of 43) and 8 gestures (out of 16) for training. The S1 training utilizes the MyoBM-Net extractor module followed by two different classifier modules: 1) biometric (BM) classifier and 2) gesture recognition (GR) classifier. The weights of all the layers are initialized uniformly prior to the training. Consequently, the BM and GR classifiers have an output size of 35 and 8, respectively, followed by a SoftMax activation. The training objective is to minimize a dual loss function:

$$\text{Dual loss} = \gamma * \text{BM loss} + \delta * \text{GR loss} \quad (7.2)$$

where $\gamma, \delta \in [-1, 1]$ are parameters that can be varied to control the biometric/gesture recognition learning of the network ($\gamma, \delta = \{1, 1\}$ for this study). In each episode, the extractor and classifier modules are trained for 10 epochs each, and the network is continuously trained for 25 such episodes. The trained weights of the MyoBM-Net (extractor module) are further utilized for the S2 training.

S2 Training

The objective of the S2 training is to extract features that are specific to a particular user and their authorizing code sequence. A trained S2 feature extractor processes the input signals and provides the authorization/rejection decision. The S2 training data includes: 1) the true user's data from the code sequence gestures ($n = 4$) from the training day (Day 1/2 for SCD and Day 1,2 for CCD), and 2) some false users' ($n = 10$) data of the same code sequence from the training day. The S2 training involves a binary classification (true/false), and hence training data from the two classes is needed to train the network architecture. For the S2 training, the MyoBM-Net and a classifier module are used. The weights from MyoBM-Net layers are transferred from the S1-trained network, while the weights of the classifier module are initialized uniformly. The training

objective is to minimize the $BM\ loss$ of the classifier network. The network for each code is trained for 30 epochs and the trained S2 network is used for biometric matching. Therefore, for a code-sequence of length four, an S2-trained MyoBM-Net comprises four separately trained feature extractors (one for each code).

To understand the effect of S1 training on authentication performance, a baseline condition (S2-BL) is added to the analysis, where S1 training is bypassed and the layers of the S2 network are initialized with uniform weights. Similar to the S2 training, S2-BL is trained for 30 epochs. Therefore, to investigate the effect of training paradigms, the three conditions (S1, S2-BL, and S2) are analyzed separately and the results are compared. The networks were trained in Python using the PyTorch framework version 2.0.1. Computation was performed on an Intel(R) Core (TM) i7-6700K CPU operating at 4.00 GHz and an NVIDIA GeForce GTX 1070 GPU equipped with 8GB of memory.

7.2.5 Performance Evaluation

After a feature extractor network is trained, the authentication task involves feature extraction, enrollment of extracted features, biometric matching, and authentication decision, followed by performance evaluation, as explained in this section and illustrated in [Figure 7.3 B](#). Also, the different biometric authentication scenarios and the analysis involved in comparing different networks are discussed further.

Biometric matching

After network training, the training data (Day 1/2 for SCD and Day 1, 2 for CCD) from a specific user's code sequence is used as input for the feature extractor network. Relevant features are extracted which serve as the enrollment data for an authenticating user. Therefore, for a specific user j and a code i , the biometric template is defined as

$$B_{i,j} = \left[\mu_{i,j}, \sum_{i,j} \right], \quad i \in \{1, 2, 3 \dots C\} \quad (7.3)$$

where $\mu_{i,j}$, $\sum_{i,j}$ are the mean vector and covariance matrix, respectively, and C is the code-length ($C = 4$ for the study).

In the testing phase, the presenting user provides their input data, followed by feature extraction using the identical network that was used for the enrollment data. The input consists

of feature vectors from every code, denoted by p_i , its matching score, $S_{i,j}$, was defined as the Mahalanobis distance between the sample and the class centroid ($\mu_{i,j}$, $\Sigma_{i,j}$):

$$S_{i,j}(p_i) = \sqrt{(p_i - \mu_{i,j})^\top \Sigma_{i,j}^{-1} (p_i - \mu_{i,j})} \quad (7.4)$$

Following the matching module, a multicode framework is employed utilizing the score-level fusion [106]. A simple summation of matching scores was performed to obtain the final score (S_T) for the j^{th} user:

$$S_{T,j}(p_i) = \sum_{i=1}^C S_{i,j} \quad (7.5)$$

Therefore, for every sample of a presenting user, a final score is generated, which can be utilized to get a performance estimate. Typically, the detection error tradeoff (*DET*) curve is generated, where the false rejection rate (*FRR*) is plotted against the false acceptance rate (*FAR*) by varying the score thresholds (for accepting or rejecting the claim). A widely used metric, the equal error rate (*EER*) is that point on the *DET* curve where *FAR* is equal to *FRR*. The lower the *EER* value the better the verification performance. To generate the *EER* value, scores from the true user and imposters are utilized as explained in the next subsection.

Testing scenarios:

Consistent with the previous multi-day study, two biometric authentication scenarios are assessed: 1) normal test and 2) leaked test [108]. A normal test is a scenario where the correct code sequence was only known to the true user, but the imposter had no knowledge of the code sequence. A random code sequence from the imposter was generated, which corresponds to a ‘guess’ made by the imposter to attack the true user’s authentication. A leaked test is a scenario where the code sequence is compromised, and hence it is known both to the true user and the imposter. The imposter simulates presenting the —correct code sequence to attack the true user’s authentication.

As the database includes 43 users, the authentication analysis is repeated for every user. In each iteration, for every true user, there can be 42 imposters. It was previously mentioned that out of the 42 users, the network training involved false-class data from 10 randomly chosen users. This implies there are 32 unknown users that would qualify as imposters, and hence, for obtaining imposter scores. For the true scores, only the true user’s scores were considered. The true scores and imposter scores were pooled together to determine the *EER* for that specific true user. These analyses were repeated 30 times for each randomly generated code sequence of length four (#iterations = 43×30).

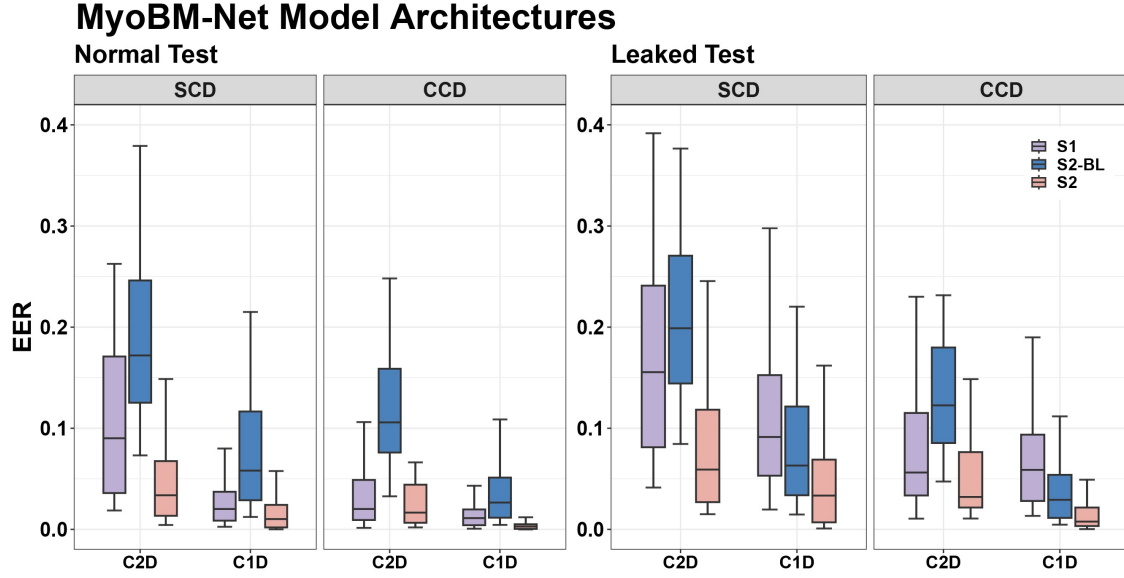


Figure 7.4: Authentication performance comparison of network architectures and training paradigms. For the normal test (left) and leaked test (right), the *EER* values for SCD and CCD training are plotted. The x-axis represents the two network architectures: C2D and C1D, and the y-axis represents the performance error. The colors represent the three training paradigms: S1, S2-BL and S2. Each boxplot represents the interquartile range (IQR, 25th – 75th percentile) and the center horizontal line represents the median *EER* value. The whiskers (solid vertical lines) represent the datapoints within the $1.5 \times \text{IQR}$ threshold.

Statistical analysis

The two cross-day training conditions investigated in the study were SCD and CCD. The median *EER* value for each analysis was reported along with IQR (first quartile-third quartile). The two testing scenarios (normal test and leaked test) were investigated separately. For each of the scenarios, the network architectures (C2D and C1D) and the training paradigms (S2-BL, S1, and S2) were investigated separately. A Kruskal-Wallis (KW) test was used to determine if there was a significant difference between network architectures, training paradigms, testing scenarios, and the two cross-day analyses. In case of a significant difference, a pairwise Wilcoxon rank-sum test was performed for post-hoc analysis. Further, the *EER* of S2 training is compared with the traditional feature extraction method. A KW test and post-hoc analysis are performed to determine if there is a significant difference between the feature extraction methods (FDT, C2D,

and C1D). All statistical tests were performed using RStudio 1.0.136 (RStudio, Boston, MA).

Feature space analysis

Feature space analysis was performed to examine how well users can be separated from each other. The t-distributed stochastic neighbor embedding (t-SNE) tool was employed to visualize the feature space for all 43 users in a two-dimensional (2D) space [126]. In our study, the t-SNE feature spaces were analyzed after each of the S1 and S2 training stages. For simplicity, CCD training was only considered. In the S1 t-SNE plots, the features of all 43 users from the testing day were extracted using the S1-trained MyoBM-Net network. In the S2 t-SNE plots, the features of each user from the testing day were extracted using the S2-trained MyoBM-Net network. The visualization compared the testing day features of the true user and impostor to the enrollment features of the true user.

The feature space can be further quantified using the Davies-Bouldin index (DBI), which is a measure of how well each cluster is separable from each other. DBI is evaluated by obtaining the ratio of within-class similarity and between-class similarity. The DBI analysis was performed only for the S1-trained network to understand the feature space separability of different users. Let S_i, S_j represent the within-class similarity of users i and j . Let $D_{i,j}$ denote the between-class similarity between the two users. The ratio is calculated as

$$R_{i,j} = \frac{S_i + S_j}{D_{i,j}} \quad (7.6)$$

$$\text{DBI} = \frac{1}{N} \sum_{i=1}^N \max R_{i,j}, \quad N = 43 \quad (7.7)$$

For each user i , the pair with the worst-case separability ($\max R_{i,j}$) is calculated. These values are averaged for all the users to obtain the DBI value as shown in equation 7.7. The analysis was performed for both the SCD and CCD scenarios, and the DBI values were reported. Statistical analysis was performed to test for significant differences between feature extraction methods similar to Section II.E.3

7.3 Results

The result section is organized as follows. The authentication performance is evaluated and compared between the training paradigms of MyoBM-Net. The next section compares the different

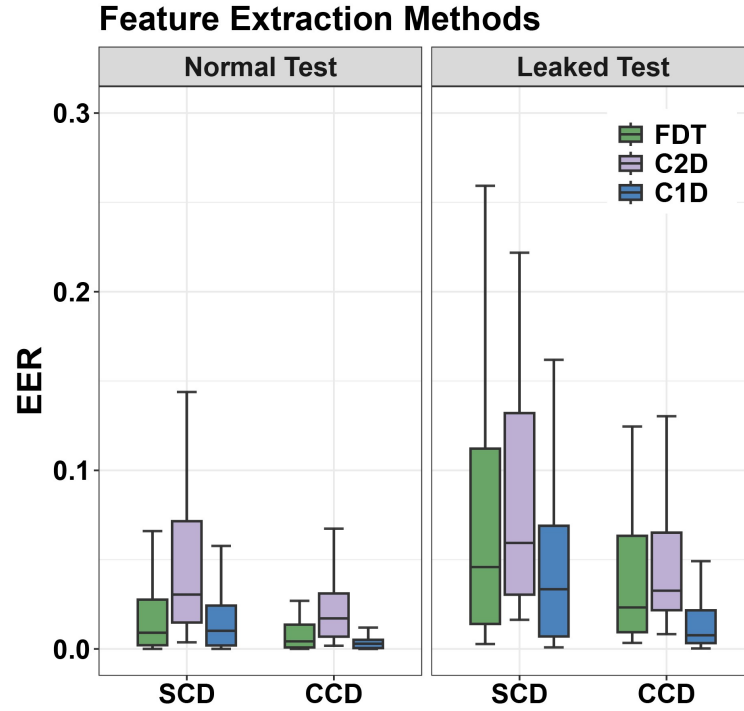


Figure 7.5: Authentication performance of S2-trained networks is compared with the traditional methods. For the normal test (left) and leaked test (right), the EER values for SCD and CCD training (x-axis) are plotted. The y-axis represents the performance error. The colors represent the feature extractors: FDT, C2D, and C1D. Box and whisker The boxes and whiskers are similar to Figure 7.4

network architectures using S2 training and further compares them to the traditional method. The following section reports the feature space analysis results of the S1 training and compares it to the traditional method.

7.3.1 Comparison of training paradigms

Figure 7.4 shows the authentication performance of different training paradigms (S1, S2-BL, and S2) and the network architectures (C2D and C1D). In the normal test scenario (Figure 7.4, left), the EER for the CCD analysis for all the network architectures and corresponding training paradigms was significantly lower than their SCD counterparts. Similarly, in the leaked test

scenario, the *EER* for the CCD analysis was significantly ($p < 0.001$) lower than those for the SCD analysis. It was also observed that, for all the network architectures and training-testing scenarios, the S2-BL had a significantly higher *EER* ($p < 0.001$) than the S2 and S1 training paradigms.

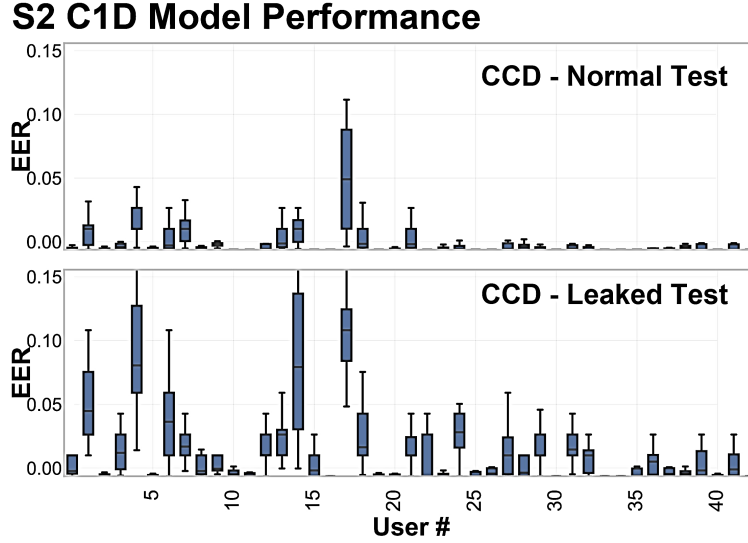


Figure 7.6: Authentication performance of S2-trained C1D network in a CCD scenario for all the users. The *EER* values (y-axis) for the normal test (top) and leaked test (bottom) are plotted. The y-axis represents the 43 users in the database. The boxes and whiskers are presented similar to [Figure 7.4](#)

Specifically, for the C2D network, the S2 training in a normal test-SCD scenario ([Figure 7.4](#), left) had a significantly lower ($p < 0.001$) *EER* (Med 0.030, IQR 0.014–0.071) than the corresponding S1 training (Med 0.035, IQR 0.018–0.089), and the baseline S2-BL training (Med 0.171, IQR 0.125–0.246). In the normal test-CCD scenario, there were no significant differences ($p = 0.83$) between the *EER* of S2 training (Med 0.017, IQR 0.006–0.031) and S1 training (Med 0.015, IQR 0.008–0.037). In the leaked test-SCD scenario ([Figure 7.4](#), right), a similar trend followed where S2 training had a significantly lower ($p < 0.001$, Med 0.059, IQR 0.030–0.132) *EER* than the other training paradigms. In the leaked test-CCD scenario, the S2 training (Med 0.021, IQR 0.041–0.065) had a significantly lower ($p < 0.001$) than the other training paradigms.

For the C1D network (Fig 4), the *EER* of S2 training in a normal test-SCD scenario (Med 0.010, IQR 0.002–0.024) was significantly lower ($p = 0.048$) than the S1 training (Med 0.020, IQR 0.001–0.037). For the normal test-CCD scenario, the *EER* of S2 training (Med 0.003, IQR 0.001–0.005) was significantly lower ($p < 0.001$) than S1 training (Med 0.011, IQR 0.004–0.019). For

the leaked test-SCD (Figure 7.4, right), the *EER* of S2 training (Med 0.033, IQR 0.006–0.069) was significantly lower ($p < 0.001$) than S1 training (Med 0.091, IQR 0.053–0.152). For the leaked test-CCD, the *EER* of S2 training (Med 0.008, IQR 0.003–0.021) was significantly lower ($p < 0.001$) than S1 training (Med 0.058, IQR 0.028–0.093). A detailed breakdown of all the users' performance for the C1D network is provided in Figure 7.6, where it was observed that the majority of the users had a low *EER* value (< 0.05).

7.3.2 Comparison of feature extraction methods

From Figure 7.4, it was observed that the *EER* of C2D was significantly higher ($p < 0.001$) than C1D, in all the test scenarios and the training paradigms. Figure 7.5 shows the authentication performance of different extraction methods (FDT, C2D, C1D). For the convolution networks, the S2 training was used for comparison. The FDT and C1D had lower *EERs* in all the scenarios as further described. In the normal test-SCD scenario, the FDT had the lowest *EER* (Med 0.009, IQR 0.002–0.027), however, it was not significantly different from the C1D ($p = 0.86$, Med 0.010, IQR 0.002–0.024). In the normal test-CCD scenario, the C1D had the lowest *EER* (Med 0.003, IQR 0.001–0.005), however, it was not significantly different from the FDT ($p = 0.47$, Med 0.004, IQR 0.001–0.013). In the leaked test-SCD scenario, the C1D (Med 0.033, IQR 0.006–0.069) had a significantly lower *EER* than FDT ($p = 0.009$, Med 0.045, IQR 0.014–0.112). In the leaked test-CCD scenario, the C1D (Med 0.008, IQR 0.003–0.021) had a significantly lower *EER* than FDT ($p < 0.001$, Med 0.023, IQR 0.009–0.063). This indicates that the C1D had the lowest *EER* in all the testing scenarios. The low *EER* values of the C1D are further demonstrated for all users in Figure 7.6.

7.3.3 S1 and S2 feature space analysis

Figure 7.7 shows sample t-SNE plots (left) and the DBI values (right) of the S1-trained networks and their comparison to the traditional method. In the SCD scenario, it was observed that DBI values of C1D (Med 1.83, IQR 1.72–1.94) were significantly lower ($p < 0.001$) than the other feature extraction methods. Similarly, in the CCD scenario, the DBI values of C1D (Med 1.87, IQR 1.68–1.91) were significantly lower ($p < 0.001$) than the FDT (Med 2.38, IQR 2.12–2.66) and C2D (Med 2.28, IQR 2.05–2.41). Two-dimensional t-SNE plots were also examined to understand the separability of different users in the feature space. Figure 7.7 (left) shows a sample t-SNE plot in a CCD scenario comparing the FDT, C2D, and C1D feature spaces for a specific code sequence. It was observed that the C1D had more distinct clusters compared to the FDT, and C2D, which was also confirmed with the DBI values. In Figure 7.8, the cross-day

S2 Feature Space Analysis

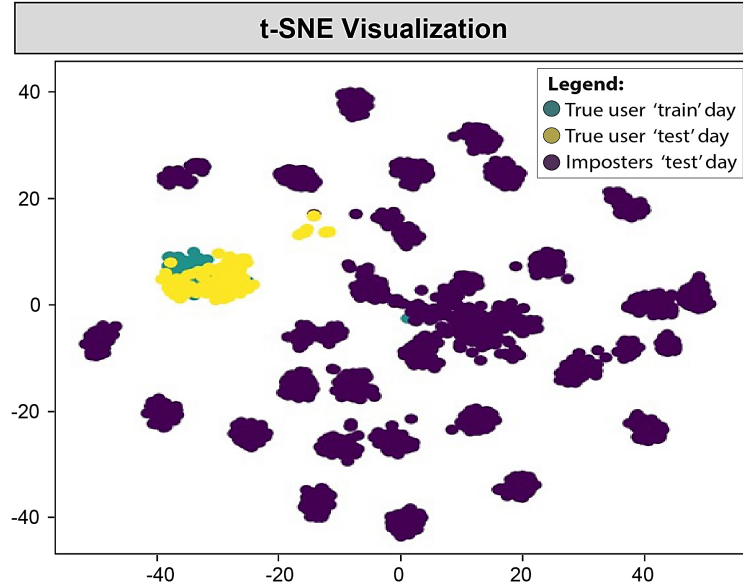


Figure 7.7: The feature space analysis of S1-trained networks. The t-SNE plots (1-3, from left) and DBI (right) values are plotted. The x-axis and yaxis for the t-SNE plots represent the reduced dimensions. The plots show the clustering of users' feature space for an identical code sequence in the three feature extraction methods (FDT, C2D and C1D). In the DBI plot (right), the x-axis denotes the training scenarios, and the y-axis denotes the DBI values. The colors represent the feature extractors: FDT, C2D, and C1D. The boxes are presented similar to [Figure 7.4](#)

authentication performance is visualized for a sample user in the C1D network in a CCD-leaked test scenario. It was observed that the features from the S2-trained C1D network are clustered together and are well separated from the imposter users. This demonstrates the capability for superior authentication performance.

7.4 Discussion

This study presented cross-day authentication performance of MyoBM-Net convolutional feature extraction. Two variations of network architectures and three training paradigms were investigated to obtain an accurate, robust, and trainable feature extractor for authentication using wrist-EMG. Additionally, two training conditions (SCD and CCD) and two testing conditions

S1 Feature Space Analysis

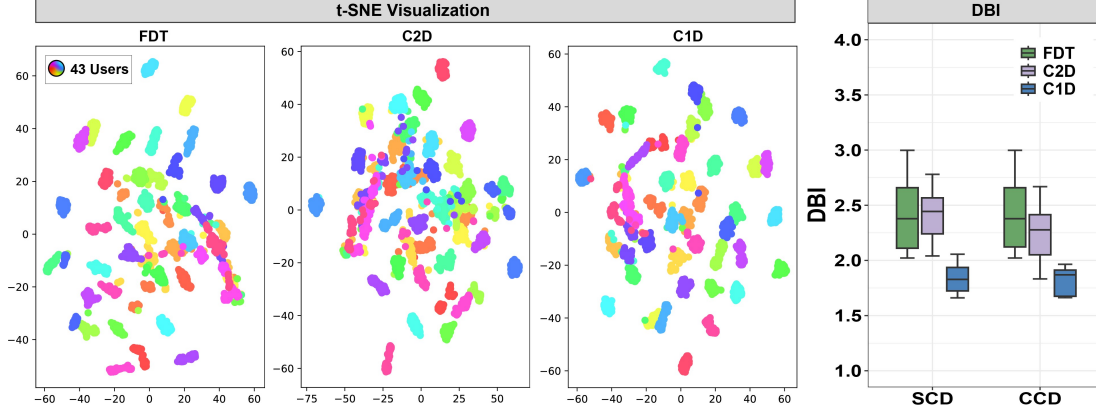


Figure 7.8: The feature space of S2 trained C1D network in a CCD testing for a sample true user. The t-SNE plot shows the visualization of feature space of a true user on two different days and their relative separation from the imposters.

(normal test and leaked test) resulted in four practical scenarios for performance evaluation. As many results have similar trends, the results of the CCD analysis are discussed in the following sections.

7.4.1 Authentication performance analysis

CNN architectures

The two MyoBM-Net networks (C2D and C1D) have similar network parameters (as shown in Table 7.1) and only vary in the use of spatial information along the channel axis. It was observed that among the two S2 networks, the C1D had better performance in the normal test-CCD (EER 0.003), leaked test-CCD (EER 0.008), normal test-SCD (EER 0.010) and leaked test-SCD (EER 0.033). Overall, the performance of C1D surpassed that of C2D. The C1D architecture utilizes a kernel size of $C \times k$, where $C = 6$, and k is the filter size on the time-axis. C2D, on the other hand, uses $1 \times k$ filters which do not utilize any spatial information. The results contradicted a previous gesture recognition (GR) study which recommended the use of channel-wise convolutions rather than spatial information [43]. The difference in findings might be attributed to the unique spatial correlation between channels in users (biometrics) which hinders a generalized GR model.

CNN training paradigms

The authentication performance of the three training paradigms (S1, S2-BL, and S2) was compared for the different network architectures. The results showed that the S2 training had a lower *EER* than S1 and S2-BL training, for all the network architectures. This suggests that, for authentication applications, a trainable, user-specific feature extractor (S2) is more accurate than a generalized one (S1). Additionally, it was found that the baseline condition (S2-BL) resulted in the highest *EER* among the training paradigms, suggesting S2-training performance is worse if S1 training is bypassed. Therefore, it is important to perform the S1 training and fine-tune the network for a specific user with S2 training.

Another study utilized a two-step training strategy where in step one, the gesture is recognized, and, in step two, the correct gesture is used for personal identification [74]. Two different CNN-LSTM networks are used for both stages. Comparatively, the proposed two-stage training strategy utilizes both gesture learning and biometrics in a single stage. Additionally, the same network is used for both stages, which not only reduces computational overhead but also retains a balance between user-specific and gesture-specific information. This is beneficial for the cancelable nature of EMG biometrics, where once the code is stolen, a new code can be used without performance loss [62, 114].

Feature space analysis

The t-SNE plots in Figure 7.7 suggested better clustering in 2D spaces for the MyoBM-Net networks, specifically with C1D. For the separability of users in the feature space, C1D had the lowest DBI values (DBI 1.83) which was significantly lower than the FDT feature extraction (DBI 2.37). This suggests samples from a specific user are clustered together and are more separable from other users. A low DBI indicates a higher potential of MyoBM-Net for personal identification applications in future studies. In a sample plot (Figure 7.8), it was observed that the samples from a single user on two different days are clustered together. This suggests that the network learns cross-day features which provide a more robust authentication performance.

Comparison with traditional methods

The FDT feature extraction resulted in a median *EER* of 0.004 in the normal test-CCD and 0.017 in the leaked test-CCD. This was lower than the median *EER* (0.039 and 0.068, respectively) reported in a similar cross-day analysis that utilized the FDT feature extraction [108]. The difference in *EER* values was due to the differences in SCD and CCD training-testing data, where

previously cross-validation was used, and in the current study training data was limited to Day 1/2, testing on Day 3, and the number of trials were reduced. The present study provides more consistent and practical testing conditions, and hence, the comparison between MyoBM-Net and traditional methods was more meticulous.

Compared to FDT, C1D had a higher performance in three of the four train-test scenarios. The FDT feature extraction is similar to the S1-trained MyoBM-Net, due to the pre-trained weights and one-size-fits-all nature. As FDT does not incorporate trainable weights in their computation, obtaining a lower *EER* is not attainable. The MyoBM-Net, on the other hand, allows finetuning to specific users, thus able to attain even lower *EER* with the S2 training, compared to the S1 training. This demonstrates the benefit of utilizing trainable weights in convolutional feature extraction methods. Overall, the results reflect the superiority of the C1D network for biometric authentication.

7.4.2 Comparison with state-of-the-art EMG Biometrics

The high biometric performances in the present study (authentication *EER* 0.003) are comparable to conventional biometrics such as fingerprints, facial recognition, and iris (*EER* range 10^{-4} - 0.20, *rank-5 accuracy* > 90%) [25]. Conventional biometrics have undergone decades of research and commercialization and are deeply integrated into our daily lives. Comparatively, most of the EMG-Biometrics studies are recent with many ongoing projects. Differences in experimental setup, protocol, and biometric analysis pose challenges for a systematic comparison of our proposed feature engineering. For comparison, we have only considered those studies that performed a cross-day train-test analysis.

One study utilized CNN for cross-day authentication and a codelength of four reported a leaked-test *EER* > 0.15 and normal-test *EER* > 0.025 [114]. Another study investigated password-based EMG biometrics of codelength eight and reported a normal-test *EER* of 0.012 and a leaked-test *EER* of 0.1496 [64]. Another study involving a similar multi-code password of codelength 12 reported a normal-test *EER* of 0.0013 and a leaked-test *EER* of 0.0273 [61]. In another cross-day investigation reported an average *EER* of 0.1634 for eight individual gestures [117], however, different codelengths were not investigated. These findings can help ongoing research on unlocking smartphones or developing wristband devices for biometric authentication [29, 126].

7.4.3 Long-term and practical viability.

In the present study, GRABMyo was chosen for the dataset, as other datasets suffer a trade-off between user sample size and the number of recording days [107]. The study explored the SCD and CCD training scenarios as well as the normal-test and leaked-test scenarios for the long-term practical viability of EMG biometrics. The results of the study showed that for the same testing data, the CCD training had a higher authentication performance than SCD. While the training loss was identical in the two scenarios, the CCD training utilizes data from two different days. This enables the extraction of only those features that are robust to multi-day variations such as electrode shifts, skin conditions, and fatigue. In a recent study, it was recommended to perform incremental training and include more training data over multiple days for robust gesture recognition performance [149]. The two testing scenarios assess the dual-security (knowledge-level and a biometrics-level), a unique property of EMG-Biometrics [106]. The normal test authentication performance was higher than the leaked test performance for all the networks, suggesting that the authentication is more secure when the impostor does not know the codes. Notwithstanding, the C1D network architecture resulted in a very low median *EER* of 0.008 for the leaked test-CCD scenario. This signifies that even when the code is compromised, the biometrics layer protects true users from imposters, thus demonstrating the two-factor authentication of EMG-Biometrics.

Therefore, in practical scenarios, trainable convolutional networks like MyoBM-Net demonstrate the high performance of EMG-based biometric authentication. As the data used for network training was low-density wrist EMG, this further boosts the development of wearable devices and smartwatches [126]. Furthermore, eight-channel wrist EMG has been utilized for gesture recognition by deploying 2D CNN networks in edge devices [66]. These results bolster the commercialization and adoption of EMG-based authentication in daily use.

7.5 Chapter Conclusion

The study presented MyoBM-Net, a convolutional feature extraction method with a dual-training strategy. The proposed training utilized a more generalized gesture-learning and biometrics training in the first stage, and a fine-tuned user-specific authentication training for the second stage. The results were tested on a large EMG-based hand gestures dataset (43 users x 3 sessions = 129 cases). The proposed method attained a very low *EER* of 0.003 for biometric authentication. When the code was leaked, the authentication performance resulted in an *EER* of 0.008. These results indicate superior performance of convolutional feature engineering, thus bringing EMG-Biometrics to par with conventional biometrics such as fingerprints and facial recognition. Thus MyoBM-Net presents an accurate, robust, and long-term viable feature extraction method for

biometrics applications using wrist EMG signals. Future studies should explore other deep feature engineering and domain adaptation techniques to improve the authentication performance of EMG-Biometrics.

Chapter 8

Conclusions

8.1 Summary of Results

This research aimed to investigate the potential of EMG-based biometrics as a novel solution to the challenges faced by conventional biometrics, such as fingerprint and facial recognition. While EMG biometrics possesses inherent characteristics such as liveness, spoof resistance, and easy accessibility, achieving robust and higher performance is essential for authentication and identification applications. A biometric system should be able to 1) differentiate the genuine user from all other users (uniqueness) 2) retain the biometric performance over multiple sessions and across multiple days (robustness). To investigate the above properties the thesis was divided into five main objectives: 1) determining the optimal EMG system parameters 2) developing a multi-code framework for combining gestures 3) data acquisition and creating a large-sample dataset with multiple days of recording, 4) testing the robustness of the above framework and EMG configuration on the recorded dataset, and 5) developing a deep convolutional feature engineering that can outperform the traditional methods.

Under objective 1, several insights were gained into the best feature extraction methods and the number of EMG channels for optimal performance of biometric applications. It was found that the TD (time-domain) features performed better than the FDT (frequency-domain) features in both the authentication and identification modes of biometrics. However, both the feature extraction methods achieved comparable performance to previous research and other biometric traits. Additionally, for all the feature extraction methods, a four-channel EMG setup had a plateau in biometric performance with the further addition of EMG channels. The finding motivates the design of an easy-to-wear armband with four equally spaced electrodes for practical

applications. Therefore, an accurate EMG-based biometric system for biometric applications can be designed using a TD feature extraction method and a four-channel EMG setup.

Under objective 2, developmental work resulted in a multi-code EMG framework, which facilitates a practical scenario where a user can use a sequence of hand gestures as an access code. Insights were gained into fusion strategies of the framework and the optimal codelength (number of hand gestures) needed for a secure biometric system. For the authentication mode, the Decision-WM (a weighted fusion of the final stage decisions) had the lowest EER, and for the identification mode, the Score-Norm (a normalized addition of individual scores) resulted in the highest recognition accuracy. The fusion of codes resulted in higher biometric performance than the single-code configuration. The recommended codelength for the two application modes was four, however, there is flexibility in choosing the desired codelength based on the subject-pool size and convenience of performing multiple gestures. Therefore, an improved multicode EMG system with dual security can be designed using specific fusion schemes and using the recommended codelength.

The previous two objectives were investigated on a sample size of 24 with only a single recording-day. This was a limitation of the findings, as EMG is known to suffer performance degradation in a multi-day application due to electrode shift, changing skin and health conditions. Under objective 3, several contributions were made to address these limitations. Data acquisition involving 43 participants over three different days (the separation over a month) was completed. The participants performed 16 hand and wrist gestures with 7 repetitions on each day (identical to the first two objectives). The data quality was analyzed and was made publicly available as the GRABMyo dataset. GRABMyo was the largest EMG-based dataset in terms of the number of participants and the recording trials (43 participants x 3 sessions = 129 recordings). This dataset will benefit other researchers in furthering the research on EMG-biometrics and hand gesture recognition.

Under objective 4, insight was gained into the robustness of wrist EMG-Biometrics over multiple recording days. A detailed multi-day analysis considering within-day and cross-day scenarios was performed using the GRABMyo dataset. Validating the findings of objective 2, the multi-code biometric framework had superior performance than the single-code system, suggesting enhanced security with customizable gesture passwords. In a cross-day scenario, the authentication mode resulted in an EER as low as 0.03. The identification mode presented a rank-5 accuracy of 93% which is comparable to most novel biometric traits. The findings can facilitate the development of wearable EMG-based bracelets and wristbands for biometric authentication and identification.

Under objective 5, insights were gained into the utilization of deep learning for an improved authentication mode over the traditional baselines set by objective 4. MyoBM-Net, a deep, con-

volutional feature extraction method was developed that incorporated a novel two-step training strategy. The cross-day performance was analyzed using the GRABMyo dataset. The proposed method attained a very low EER of 0.003 for biometric authentication. When the code was leaked, the authentication performance resulted in an EER of 0.008. The performance was significantly higher than the traditional baseline, as the deep model was trained specifically for each user. These findings indicate superior performance of convolutional feature engineering, thus bringing EMG-biometrics to par with conventional biometrics such as fingerprints and facial recognition.

Therefore, completing the thesis work made significant contributions towards achieving high biometric performance with EMG-Biometrics. These high performances help realize a novel biometric trait that is both unique and robust for multi-day applications. The findings show the potential of using EMG signals for biometric applications in real-world scenarios. The multi-code framework facilitates the use of gestures as passcodes. The large multi-day dataset will support further research on EMG-based biometrics and other gesture recognition applications. The MyoBM-Net architecture will enable the development of new applications using the GRABMyo dataset, leading to accurate and robust biometric performance. This could lead to EMG-based biometrics being used as an alternative to traditional biometric methods.

8.2 Limitations and Future Research

While the thesis accomplished the objectives, there exist certain limitations to the research. In objectives 1 and 2, the performed studies faced limitations of the subject sample size and the lack of multi-day analysis. Due to the COVID-19 conditions prevailing during the duration of the investigation of the two studies, it was impractical to record more data. Therefore, the findings could vary when tested on a cross-day scenario. To address this a large-scale data collection was performed for further research. Corroborating the findings of objective 2, the multi-code EMG-Biometrics resulted in higher performance than the single-code configuration in a cross-day scenario. However, it was observed that the FDT was more robust than the TD feature extraction, contrary to the findings of objective 1.

In objectives 3 and 4, there existed some limitations in the GRABMyo data collection. Multiple non-stationary factors such as electrode shifts and varying force levels have not been included in the experimental protocol, more detail on these factors can be found on [Table 8.1](#). Multiple shifted electrode positions and multiple force levels could benefit the GRABMyo dataset, as the effect of these two factors on the multi-day performance can be estimated. However, this would lead to a considerable increase in the number of trials for every participant and hence extensively prolong the overall data collection. Another limitation is the dataset involves static hand

Table 8.1: Factors affecting EMG-Biometrics performance

Factors	Description	Effects
Variation in electrode placement over multiple days.	No marks were left on the skin, therefore the electrode placement on subsequent days had some variations.	The shift in positions for a particular channel can result in differences in EMG signal amplitude and frequency and hence affect the EMG feature pattern overall, thus affecting biometric performance.
Variation in force levels over multiple trials and days.	Participants were instructed to perform the gestures as they would do normally in their daily life. As no feedback of force level was provided, there might be varying force levels across participants as well within trials for the same participant.	The varying force levels can result in variation of EMG features which can affect the biometric performance.
Fatigue	The participants perform gestures over multiple trials in a single session and can feel some fatigue at the end. The order of the gestures was randomized to reduce the effect of fatigue.	Fatigue results in variation of EMG features which can affect the biometric performance.
Surface preparation	Some participants have dry or sweaty skin surface. The skin was thoroughly prepared for each participant by cleaning and abrasion and the gel electrodes were placed.	The varying skin conditions (dryness, sweat) can affect the signal quality and hence the biometric performance.
Gel electrodes vs Metal plate electrodes	The study utilized gel electrodes and presented results in an experimental setting. Practical scenarios would utilize more feasible metal-plate electrodes.	The gel electrodes result in higher signal quality which can result in higher biometric performance. In a practical scenario using metal-plate electrodes, the results might be affected due to lower signal quality.

and wrist gestures. In a practical scenario, human movements are dynamic. This limitation can be mitigated as the biometric system can process the signals with a slight delay (1 sec) which will facilitate post-processing to extract static components from the EMG signal. As shown in the above table multiple experimental factors affecting the EMG signal, can further affect the biometric performance. Further studies should explore the factors by creating control and experimental conditions with these factors to understand the exact effect of such factors.

In objective 5, the limitations faced by the study were to replicate the analysis of objective 4 as a traditional baseline. The multi-day analysis of objective 4 followed a leave-one-trial-out cross-validation (LOOCV) nature, where iteratively one day and one trial were used for testing and the remaining trials were used for training, and this process was repeated. However, the convolutional networks suffered from increased time in model training, and hence the computational time of all the LOOCV runs was enormous. To mitigate this, the cross-day iterations were reduced to where day 3 was only used for testing and day 1 and day 2 for training. Additionally, the training data were randomly chosen from 5 out of 7 repetitions of a gesture. This analysis was consistent for both the convolutional feature engineering and the traditional FDT feature engineering.

While the MyoBM-Net utilized the convolutional networks, currently other deep architecture comprising LSTM, GRU, transformers and a combination need to be explored and compared. With other data-types such as images and time-series, these networks provided accurate label-classifications. One study using CNN-LSTM network has shown improved biometric performance while using surface EMG [74]. Therefore, further research is warranted using other network types. While the current research uses a comprehensive dataset in terms of number of users and days, the studies utilize low-density or sparse EMG channels. Other studies on EMG-Biometrics have used high-density (HD) EMG channels. The accurate performance using low-density EMG is advantageous as it reduces computation cost and has higher feasibility of use. However, a study comparing the biometric performance of low-density and HD EMG for similar experimental factors is yet to be performed. The analysis should investigate the computational complexity in terms of training time, number of floating operations (FLOPs), number of trainable parameters and inference time.

Despite the large nature of GRABMyo dataset, further work is needed to extend the dataset with even more recording days. A different study investigated EEG biometric characteristics such as permanence and uniqueness by analyzing six recording sessions spanning over 3 years [88]. The inclusion of transfer learning in EMG biometric analysis can facilitate such a longitudinal investigation. The convolutional feature engineering facilitates research into transfer learning approaches that are insensitive to electrode shift, and hence, beneficial for accurate biometric authentication [112]. The 6-channel electrode setup was part of a bigger dataset, *i.e.*, GRAB-Myo, which includes 28 channels recorded from the forearm and wrist region [56, 107]. While

preliminary results suggested no difference between the forearm and the wrist setup, further investigations could lead to novel findings [104]. Other strategies to improve the authentication and identification performance can be further investigated by implementing different fusion strategies [106].

8.3 Practical Significance

There have been ongoing warnings of cyber-attacks that could pose security risks at the Canadian border [131]. A data leak in the UK exposed the fingerprints of 1 million individuals [40]. In addition to the economic loss, worth billions of dollars, such identity theft attacks can have a lifetime impact on individuals, restricting them from using current biometric techniques. Unlike fingerprints, once the EMG-Biometrics is exposed, the user can change the gesture codes to prevent further hacking incidents. The high-accuracy and multi-day robustness of EMG-Biometrics provide a more secure solution to the limitations of present-day biometrics. While most current securities involve a combination of password and biometrics (multi-layer), EMG-Biometrics could inherently facilitate dual-security with easy-to-perform hand gestures and their associated signals. Findings of this doctoral research demonstrate the potential of EMG-Biometrics as a novel biometric trait and facilitates the development of wearable wristbands for practical applications.

References

- [1] Foteini Agrafioti, Francis M Bui, and Dimitrios Hatzinakos. Secure telemedicine: Biometrics for remote and continuous patient verification. *Journal of Computer Networks and Communications*, 2012, 2012.
- [2] Christoph Amma, Thomas Krings, Jonas Böer, and Tanja Schultz. Advancing muscle-computer interfaces with high-density electromyography. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 929–938.
- [3] Ali Raza Asif, Asim Waris, Syed Omer Gilani, Mohsin Jamil, Hassan Ashraf, Muhammad Shafique, and Imran Khan Niazi. Performance evaluation of convolutional neural network for hand gesture recognition using emg. *Sensors*, 20(6):1642, 2020.
- [4] Manfredo Atzori, Arjan Gijsberts, Claudio Castellini, Barbara Caputo, Anne-Gabrielle Mittaz Hager, Simone Elsig, Giorgio Giatsidis, Franco Bassetto, and Henning Müller. Electromyography data for non-invasive naturally-controlled robotic hand prostheses. *Scientific data*, 1(1):1–13, 2014.
- [5] Simone Benatti, Elisabetta Farella, Emanuele Gruppioni, and Luca Benini. Analysis of robust implementation of an emg pattern recognition based control. In *BIOSIGNALS*, pages 45–54.
- [6] Martin Bilodeau, Marco Cincera, A Bertrand Arsenault, and Denis Gravel. Normality and stationarity of emg signals of elbow flexor muscles during ramp and step isometric. *Journal of Electromyography and Kinesiology*, 7(2):87–96, 1997.
- [7] Ruud M Bolle, Jonathan H Connell, Sharath Pankanti, Nalini K Ratha, and Andrew W Senior. *Guide to biometrics*. Springer Science Business Media, 2013.
- [8] Fady S Botros, Angkoon Phinyomark, and Erik Scheme. Electromyography-based gesture recognition: Is it time to change focus from the forearm to the wrist? *IEEE Transactions on Industrial Informatics*, 18(1):174–184, 2020.

- [9] Yeong-Hyeon Byeon and Keun-Chang Kwak. Individual identification by late information fusion of emgcnn and emglstm from electromyogram signals. *Sensors*, 22(18):6770, 2022.
- [10] Evan Campbell, Angkoon Phinyomark, and Erik Scheme. Deep cross-user models reduce the training burden in myoelectric control. *Frontiers in Neuroscience*, 15, 2021.
- [11] Patrizio Campisi and Daria La Rocca. Brain waves for automatic biometric-based user recognition. *IEEE Transactions on Information Forensics and Security*, 9(5):782–800, 2014.
- [12] Qiong Cao, Li Shen, Weidi Xie, Omkar M Parkhi, and Andrew Zisserman. Vggface2: A dataset for recognising faces across pose and age. In *2018 13th IEEE international conference on automatic face gesture recognition (FG 2018)*, pages 67–74. IEEE.
- [13] Claudio Castellini, Emanuele Gruppioni, Angelo Davalli, and GiulioJournal of Physiology-Paris Sandini. Fine detection of grasp force and posture by amputees via surface electromyography. *Journal of Physiology-Paris*, 103(3-5):255–262, 2009.
- [14] Claudio Castellini and Patrick Van Der Smagt. Surface emg in advanced hand prosthetics. *Biological Cybernetics*, 100(1):35–47, 2009.
- [15] Adrian DC Chan and KB Englehart. Continuous classification of myoelectric signals for powered prostheses using gaussian mixture models. In *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE Cat. No. 03CH37439)*, volume 3, pages 2841–2844. IEEE.
- [16] Patrick PK Chan, Qiuxia Li, Yinfeng Fang, Linyi Xu, Kairu Li, Honghai Liu, and Daniel S Yeung. Unsupervised domain adaptation for gesture identification against electrode shift. *IEEE Transactions on Human-Machine Systems*, 52(6):1271–1280, 2022.
- [17] Jiangcheng Chen, Sheng Bi, George Zhang, and Guangzhong Cao. High-density surface emg-based gesture recognition using a 3d convolutional neural network. *Sensors*, 20(4):1201, 2020.
- [18] Lin Chen, Jianting Fu, Yuheng Wu, Haochen Li, and Bin Zheng. Hand gesture recognition using compact cnn via surface electromyography signals. *Sensors*, 20(3):672, 2020.
- [19] Alexander Kenneth Clarke, Seyed Farokh Atashzar, Alessandro Del Vecchio, Deren Barsakcioglu, Silvia Muceli, Paul Bentley, Filip Urh, Ales Holobar, and Dario Farina. Deep learning for robust decomposition of high-density surface emg signals. *IEEE Transactions on Biomedical Engineering*, 68(2):526–534, 2020.

- [20] Guglielmo Cola, Alessio Vecchio, and Marco Avvenuti. Continuous authentication through gait analysis on a wrist-worn device. *Pervasive and Mobile Computing*, 78:101483, 2021.
- [21] Ulysse Côté-Allard, Evan Campbell, Angkoon Phinyomark, François Laviolette, Benoit Gosselin, and Erik Scheme. Interpreting deep learning features for myoelectric control: A comparison with handcrafted features. *Frontiers in Bioengineering and Biotechnology*, 8:158, 2020.
- [22] Ulysse Côté-Allard, Cheikh Latyr Fall, Alexandre Drouin, Alexandre Campeau-Lecours, Clément Gosselin, Kyrre Glette, François Laviolette, and Benoit Gosselin. Deep learning for electromyographic hand gesture signal classification using transfer learning. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 27(4):760–771, 2019.
- [23] Ulysse Côté-Allard, Gabriel Gagnon-Turcotte, Angkoon Phinyomark, Kyrre Glette, Erik Scheme, François Laviolette, and Benoit Gosselin. A transferable adaptive domain adversarial neural network for virtual reality augmented emg-based gesture recognition. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 29:546–555, 2021.
- [24] Ulysse Côté-Allard, Gabriel Gagnon-Turcotte, Angkoon Phinyomark, Kyrre Glette, Erik J Scheme, François Laviolette, and Benoit Gosselin. Unsupervised domain adversarial self-calibration for electromyography-based gesture recognition. *IEEE Access*, 8:177941–177955, 2020.
- [25] Gabriel Dahia, Leone Jesus, and Mauricio Pamplona Segundo. Continuous authentication using biometrics: An advanced review. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 10(4):e1365, 2020.
- [26] Heather Daley, Kevin Englehart, Levi Hargrove, and Usha Kuruganti. High density electromyography data of normally limbed and transradial amputee subjects for multifunction prosthetic control. *Journal of Electromyography and Kinesiology*, 22(3):478–484, 2012.
- [27] Anany Dwivedi, Yongje Kwon, and Minas Liarokapis. Emg-based decoding of manipulation motions in virtual reality: Towards immersive interfaces. In *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 3296–3303. IEEE.
- [28] Kevin Englehart and Bernard Hudgins. A robust, real-time control scheme for multifunction myoelectric control. *IEEE Transactions on Biomedical Engineering*, 50(7):848–854, 2003.

- [29] Boyu Fan, Xiang Su, Jianwei Niu, and Pan Hui. Emgauth: Unlocking smartphones with emg signals. *arXiv preprint arXiv:2103.12542*, 2021.
- [30] Jiahao Fan, Xinyu Jiang, Xiangyu Liu, Xian Zhao, Xinming Ye, Chenyun Dai, Metin Akay, and Wei Chen. Cancelable hd-semg biometric identification via deep feature learning. *IEEE Journal of Biomedical and Health Informatics*, 26(4):1782–1793, 2021.
- [31] Yinfeng Fang, Xuguang Zhang, Dalin Zhou, and Honghai Liu. Improve inter-day hand gesture recognition via convolutional neural network based feature fusion. *International Journal of Humanoid Robotics*, 2020.
- [32] Huan Feng, Kassem Fawaz, and Kang G Shin. Continuous authentication for voice assistants. In *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking*, pages 343–355.
- [33] Graham D Fraser, Adrian DC Chan, James R Green, and Dawn T MacIsaac. Biosignal quality analysis of surface emg using a correlation coefficient test for normality. In *2013 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, pages 196–200. IEEE.
- [34] Tingting Fu, Ashirbad Pradhan, Jiayuan He, Chaoming He, and Ning Jiang. Comparison of wrist and forearm emg for multi-day biometric authentication. In *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 1–4. IEEE, 2023.
- [35] Weidong Geng, Yu Du, Wenguang Jin, Wentao Wei, Yu Hu, and Jiajun Li. Gesture recognition by instantaneous surface emg images. *Scientific reports*, 6(1):36571, 2016.
- [36] Francesca Giordaniello, Matteo Cognolato, Mara Graziani, Arjan Gijssberts, Valentina Gregori, Gianluca Saetta, Anne-Gabrielle Mittaz Hager, Cesare Tiengo, Franco Bassetto, and Peter Brugger. Megane pro: myo-electricity, visual and gaze tracking data acquisitions to improve hand prosthetics. In *2017 International Conference on Rehabilitation Robotics (ICORR)*, pages 1148–1153. IEEE.
- [37] Ary L Goldberger, Luis AN Amaral, Leon Glass, Jeffrey M Hausdorff, Plamen Ch Ivanov, Roger G Mark, Joseph E Mietus, George B Moody, Chung-Kang Peng, and H Eugene Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. 101(23):e215–e220, 2000.

- [38] Daniel Graupe, Javad Salahi, and Kate H Kohn. Multifunctional prosthesis and orthosis control via microcomputer identification of temporal pattern differences in single-site myoelectric signals. *Journal of Biomedical Engineering*, 4(1):17–22, 1982.
- [39] Yikun Gu, Dapeng Yang, Qi Huang, Wei Yang, and Hong Liu. Robust emg pattern recognition in the presence of confounding factors: features, classifiers and adaptive learning. *Expert Systems with Applications*, 96:208–217, 2018.
- [40] The Guardian. ”major breach found in biometrics system used by banks, uk police and defence firms.”, 2019.
- [41] Qiong Gui, Maria V Ruiz-Blondet, Sarah Laszlo, and Zhanpeng Jin. A survey on brain biometrics. *ACM Computing Surveys*, 51(6):1–38, 2019.
- [42] Husamuldeen K Hameed, Wan ZW Hassan, Suhaidi Shafie, Siti Anom Ahmad, Haslina Jaafar, Liyana NI Mat, and Yasir Alkubaisi. Identifying the best forearm muscle to control soft robotic glove system by using a single semg channel. In *2020 Advances in Science and Engineering Technology International Conferences (ASET)*, pages 1–4. IEEE.
- [43] Sirong Hao, Ruxin Wang, Yishan Wang, and Ye Li. A spatial attention based convolutional neural network for gesture recognition with hd-semg signals. In *2020 IEEE International Conference on E-health Networking, Application Services (HEALTHCOM)*, pages 1–6. IEEE.
- [44] Jiayuan He and Ning Jiang. Biometric from surface electromyogram (semg): Feasibility of user verification and identification based on gesture recognition. *Frontiers in Bioengineering and Biotechnology*, 8:58, 2020.
- [45] Jiayuan He, Xinyue Niu, Penghui Zhao, Chuang Lin, and Ning Jiang. From forearm to wrist: Deep learning for surface electromyography-based gesture recognition. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2023.
- [46] Jiayuan He, Dingguo Zhang, Ning Jiang, Xinjun Sheng, Dario Farina, and Xiangyang Zhu. User adaptation in long-term, open-loop myoelectric training: implications for emg pattern recognition in prosthesis control. *Journal of Neural Engineering*, 12(4):046005, 2015.
- [47] Jiayuan He, Dingguo Zhang, Xinjun Sheng, Shunchong Li, and Xiangyang Zhu. Invariant surface emg feature against varying contraction level for myoelectric control based on muscle coordination. *IEEE Journal of Biomedical and Health Informatics*, 19(3):874–882, 2014.

- [48] Jiayuan He, Dingguo Zhang, Xinjun Sheng, Jianjun Meng, and Xiangyang Zhu. Improved discrete fourier transform based spectral feature for surface electromyogram signal classification. In *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 6897–6900. IEEE.
- [49] Qin Hu, Alireza Sarmadi, Paras Gulati, Prashanth Krishnamurthy, Farshad Khorrami, and S Farokh Atashzar. X-myonet: Biometric identification using deep processing of dynamic surface electromyography. *IEEE Transactions on Instrumentation and Measurement*, 2024.
- [50] He Huang, Ping Zhou, Guanglin Li, and Todd A Kuiken. An analysis of emg electrode configuration for targeted muscle reinnervation based neural machine interface. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 16(1):37–45, 2008.
- [51] Yonghong Huang, Kevin B Englehart, Bernard Hudgins, and Adrian DC Chan. A gaussian mixture model based classification scheme for myoelectric control of powered upper limb prostheses. *IEEE Transactions on Biomedical Engineering*, 52(11):1801–1811, 2005.
- [52] Bernard Hudgins, Philip Parker, and Robert N Scott. A new strategy for multifunction myoelectric control. *IEEE Transactions on Biomedical Engineering*, 40(1):82–94, 1993.
- [53] Nisheena V Iqbal and Kamalraj Subramaniam. Robust feature sets for contraction level invariant control of upper limb myoelectric prosthesis. *Biomedical Signal Processing and Control*, 51:90–96, 2019.
- [54] Ning Jiang, Chen Chen, Jiayuan He, Jianjun Meng, Lizhi Pan, Shiyong Su, and Xiangyang Zhu. Bio-robotics research for non-invasive myoelectric neural interfaces for upper-limb prosthetic control: A 10-year perspective review. *National Science Review*, 10(5):nwad048, 2023.
- [55] Ning Jiang, Strahinja Dosen, Klaus-Robert Muller, and Dario Farina. Myoelectric control of artificial limbs—is there a need to change focus?[in the spotlight]. *IEEE Signal Processing Magazine*, 29(5):152–150, 2012.
- [56] Ning Jiang, Ashirbad Pradhan, and Jiayuan He. <https://iee-dataport.org/documents/gesture-recognition-and-biometrics-electromyography-grabmyo-dataset>, 2022.
- [57] Ning Jiang, Ashirbad Pradhan, and Jiayuan He. <https://physionet.org/content/grabmyo/1.1.0/>, 2022.

- [58] Shuo Jiang, Bo Lv, Weichao Guo, Chao Zhang, Haitao Wang, Xinjun Sheng, and Peter B Shull. Feasibility of wrist-worn, real-time hand, and surface gesture recognition via semg and imu sensing. *IEEE Transactions on Industrial Informatics*, 14(8):3376–3385, 2017.
- [59] Xinyu Jiang, Xiangyu Liu, Jiahao Fan, Xinming Ye, Chenyun Dai, Edward A Clancy, Metin Akay, and Wei Chen. Open access dataset, toolbox and benchmark processing results of high-density surface electromyogram recordings. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 29:1035–1046, 2021.
- [60] Xinyu Jiang, Xiangyu Liu, Jiahao Fan, Xinming Ye, Chenyun Dai, Edward A Clancy, and Wei Chen. Measuring neuromuscular electrophysiological activities to decode hd-semg biometrics for cross-application discrepant personal identification with unknown identities. *IEEE Transactions on Instrumentation and Measurement*, 2022.
- [61] Xinyu Jiang, Xiangyu Liu, Jiahao Fan, Xinming Ye, Chenyun Dai, Edward A Clancy, Dario Farina, and Wei Chen. Enhancing iot security via cancelable hd-semg-based biometric authentication password, encoded by gesture. *IEEE Internet of Things Journal*, 2021.
- [62] Xinyu Jiang, Ke Xu, Xiangyu Liu, Chenyun Dai, David Clifton, Edward A Clancy, Metin Akay, and Wei Chen. Cancelable hd-semg-based biometrics for cross-application discrepant personal identification. *IEEE Journal of Biomedical and Health Informatics*, 2020.
- [63] Xinyu Jiang, Ke Xu, Xiangyu Liu, Chenyun Dai, David A Clifton, Edward A Clancy, Metin Akay, and Wei Chen. Cancelable hd-semg-based biometrics for cross-application discrepant personal identification. *IEEE Journal of Biomedical and Health Informatics*, 25(4):1070–1079, 2020.
- [64] Xinyu Jiang, Ke Xu, Xiangyu Liu, Chenyun Dai, David A Clifton, Edward A Clancy, Metin Akay, and Wei Chen. Neuromuscular password-based user authentication. *IEEE Transactions on Industrial Informatics*, 17(4):2641–2652, 2020.
- [65] Xinyu Jiang, Ke Xu, Xiangyu Liu, Da Liu, Chenyun Dai, and Wei Chen. High-density surface electromyogram-based biometrics for personal identification. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*, pages 728–731. IEEE.
- [66] Hulin Jin, Zhiran Jin, Yong-Guk Kim, and Chunyang Fan. Integration of a lightweight customized 2d cnn model to an edge computing system for real-time multiple gesture recognition. *Journal of Grid Computing*, 21(4):81, 2023.

- [67] Waziha Kabir, M Omair Ahmad, and MNS Swamy. Normalization and weighting techniques based on genuine-impostor score fusion in multi-biometric systems. *IEEE Transactions on Information Forensics and Security*, 13(8):1989–2000, 2018.
- [68] Waziha Kabir, M Omair Ahmad, and MNS Swamy. A multi-biometric system based on feature and score level fusions. *IEEE Access*, 7:59437–59450, 2019.
- [69] Piotr Kaczmarek, Tomasz Mańkowski, and Jakub Tomczyński. putemg—a surface electromyography hand gesture recognition dataset. *Sensors*, 19(16):3548, 2019.
- [70] Suguru Kanoga, Takayuki Hoshino, and Hideki Asoh. Semi-supervised style transfer mapping-based framework for semg-based pattern recognition with 1-or 2-dof forearm motions. *Biomedical Signal Processing and Control*, 68:102817, 2021.
- [71] Suguru Kanoga, Takayuki Hoshino, and Hideki Asoh. Subject-transfer framework with unlabeled data based on multiple distance measures for surface electromyogram pattern recognition. *Biomedical Signal Processing and Control*, 74:103522, 2022.
- [72] Suguru Kanoga and Atsunori Kanemura. Assessing the effect of transfer learning on myoelectric control systems with three electrode positions. In *2018 IEEE International Conference on Industrial Technology (ICIT)*, pages 1478–1483. IEEE.
- [73] Rami N Khushaba. Correlation analysis of electromyogram signals for multiuser myoelectric interfaces. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 22(4):745–755, 2014.
- [74] Jin-Su Kim, Min-Gu Kim, and Sung-Bum Pan. Two-step biometrics using electromyogram signal based on convolutional neural network-long short-term memory networks. *Applied Sciences*, 11(15):6824, 2021.
- [75] Jin-Su Kim, Cheol-Ho Song, EunSang Bak, and Sung-Bum Pan. Multi-session surface electromyogram signal database for personal identification. *Sustainability*, 14(9):5739, 2022.
- [76] Agamemnon Krasoulis, Iris Kyranou, Mustapha Suphi Erden, Kianoush Nazarpour, and Sethu Vijayakumar. Improved prosthetic hand control with concurrent use of myoelectric and inertial measurements. *Journal of neuroengineering and rehabilitation*, 14(1):1–14, 2017.
- [77] Ludmila I Kuncheva. *Combining pattern classifiers: methods and algorithms*. John Wiley Sons, 2014.

- [78] Ludmila I Kuncheva, Christopher J Whitaker, Catherine A Shipp, and Robert PW Duin. Is independence good for combining classifiers? In *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000*, volume 2, pages 168–171. IEEE.
- [79] Ruggero Donida Labati, Vincenzo Piuri, Roberto Sassi, Fabio Scotti, and Gianluca Sforza. Adaptive ecg biometric recognition: a study on re-enrollment methods for qrs signals. In *2014 IEEE symposium on computational intelligence in biometrics and identity management (CIBIM)*, pages 30–37. IEEE.
- [80] Guanglin Li, Aimee E Schultz, and Todd A Kuiken. Quantifying pattern recognition—based myoelectric control of multifunctional transradial prostheses. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 18(2):185–192, 2010.
- [81] Qi Li, Anyuan Zhang, Zhenlan Li, and Yan Wu. Improvement of emg pattern recognition model performance in repeated uses by combining feature selection and incremental transfer learning. *Frontiers in Neurorobotics*, 2021.
- [82] Qingqing Li, Penghui Dong, and Jun Zheng. Enhancing the security of pattern unlock with surface emg-based biometrics. *Applied Sciences*, 10(2):541, 2020.
- [83] Qingqing Li, Zhirui Luo, and Jun Zheng. A new deep anomaly detection-based method for user authentication using multichannel surface emg signals of hand gestures. *IEEE Transactions on Instrumentation and Measurement*, 71:1–11, 2022.
- [84] Yuzhou Lin, Ramaswamy Palaniappan, Philippe De Wilde, and Ling Li. A normalisation approach improves the performance of inter-subject semg-based hand gesture recognition with a convnet. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*, pages 649–652. IEEE.
- [85] Jianwei Liu, Xinjun Sheng, Dingguo Zhang, Jiayuan He, and Xiangyang Zhu. Reduced daily recalibration of myoelectric prosthesis classifiers based on domain adaptation. *IEEE journal of biomedical and health informatics*, 20(1):166–176, 2014.
- [86] Jianwei Liu, Xinjun Sheng, Dingguo Zhang, Ning Jiang, and Xiangyang Zhu. Towards zero retraining for myoelectric control based on common model component analysis. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 24(4):444–454, 2015.
- [87] Lijing Lu, Jingna Mao, Wuqi Wang, Guangxin Ding, and Zhiwei Zhang. A study of personal recognition method based on emg signal. *IEEE Transactions on Biomedical Circuits and Systems*, 14(4):681–691, 2020.

- [88] Emanuele Maiorana and Patrizio Campisi. Longitudinal evaluation of eeg-based biometric recognition. *IEEE transactions on Information Forensics and Security*, 13(5):1123–1138, 2017.
- [89] Takamitsu Matsubara and Jun Morimoto. Bilinear modeling of emg signals to extract user-independent features for multiuser myoelectric interface. *IEEE Transactions on Biomedical Engineering*, 60(8):2205–2213, 2013.
- [90] Benjamin Miller. Vital signs of identity [biometrics]. *IEEE spectrum*, 31(2):22–30, 1994.
- [91] Md Maruf Monwar and Marina L Gavrilova. Multimodal biometric system using rank-level fusion approach. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 39(4):867–878, 2009.
- [92] Arsalan Mosenia, Susmita Sur-Kolay, Anand Raghunathan, and Niraj K Jha. Caba: Continuous authentication based on bioaura. *IEEE Transactions on Computers*, 66(5):759–772, 2016.
- [93] Ashok Muzumdar. *Powered upper limb prostheses: control, implementation and clinical application*, volume 1. Springer, 2004.
- [94] Francesco Orabona, Claudio Castellini, Barbara Caputo, Angelo Emanuele Fiorilla, and Giulio Sandini. Model adaptation with least-squares svm for adaptive hand prosthetics. In *2009 IEEE International Conference on Robotics and Automation*, pages 2897–2903. IEEE.
- [95] Mohammadreza Asghari Oskoei and Huosheng Hu. Support vector machine-based classification scheme for myoelectric control applied to upper limb. *IEEE Transactions on Biomedical Engineering*, 55(8):1956–1965, 2008.
- [96] Benjamin Paaßen, Alexander Schulz, Janne Hahne, and Barbara Hammer. Expectation maximization transfer learning and its application for bionic hand prostheses. *Neurocomputing*, 298:122–133, 2018.
- [97] Angkoon Phinyomark, Rami N. Khushaba, and Erik Scheme. Feature extraction and selection for myoelectric control based on wearable emg sensors. *Sensors*, 18(5):1615, 2018.
- [98] Angkoon Phinyomark, Pornchai Phukpattaranont, and Chusak Limsakul. Feature reduction and selection for emg signal classification. *Expert Systems with Applications*, 39(8):7420–7431, 2012.

- [99] Angkoon Phinyomark and Erik Scheme. An investigation of temporally inspired time domain features for electromyographic pattern recognition. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 5236–5240. IEEE.
- [100] João Ribeiro Pinto, Jaime S Cardoso, and André Lourenço. Evolution, current challenges, and future possibilities in ecg biometrics. *IEEE Access*, 6:34746–34776, 2018.
- [101] Matus Pleva, Štefan Korečko, Daniel Hladek, Patrick Bours, Markus Hoff Skudal, and Yuan-Fu Liao. Biometric user identification by forearm emg analysis. In *IEEE International Conference on Consumer Electronics-Taiwan*, pages 607–608. IEEE, 2022.
- [102] Ashirbad Pradhan. Frequency division technique on linear regression for robust simultaneous and proportional myoelectric control during medium and high contraction-level variation. 2018.
- [103] Ashirbad Pradhan, Jiayuan He, and Ning Jiang. Cross-day analysis of multicode surface electromyography based biometrics for personal identification. In *2023 45th Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*, pages 1–4. IEEE.
- [104] Ashirbad Pradhan, Jiayuan He, and Ning Jiang. Hand gesture recognition and biometric authentication using a multi-day dataset. In *International Conference on Intelligent Robotics and Applications*, pages 375–385. Springer.
- [105] Ashirbad Pradhan, Jiayuan He, and Ning Jiang. Performance optimization of surface electromyography based biometric sensing system for both verification and identification. *IEEE Sensors Journal*, 2021.
- [106] Ashirbad Pradhan, Jiayuan He, and Ning Jiang. Score, rank, and decision-level fusion strategies of multicode electromyogram-based verification and identification biometrics. *IEEE Journal of Biomedical and Health Informatics*, 2021.
- [107] Ashirbad Pradhan, Jiayuan He, and Ning Jiang. Multi-day dataset of forearm and wrist electromyogram for hand gesture recognition and biometrics. *Scientific Data*, 9(1):733, 2022.
- [108] Ashirbad Pradhan, Jiayuan He, Hyowon Lee, and Ning Jiang. Multi-day analysis of wrist electromyogram-based biometrics for authentication and personal identification. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 2023.

- [109] Ashirbad Pradhan, Ning Jiang, Victoria Chester, and Usha Kuruganti. Linear regression with frequency division technique for robust simultaneous and proportional myoelectric control during medium and high contraction-level variation. *Biomedical Signal Processing and Control*, 61:101984, 2020.
- [110] Ashirbad Pradhan, Usha Kuruganti, Wendy Hill, Ning Jiang, and Victoria Chester. Robust simultaneous and proportional myoelectric control scheme for individuals with transradial amputations. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 3098–3101. IEEE, 2020.
- [111] Cosima Prahm, Benjamin Paassen, Alexander Schulz, Barbara Hammer, and Oskar Aszmann. Transfer learning for rapid re-calibration of a myoelectric prosthesis after electrode shift. In *Converging Clinical and Engineering Research on Neurorehabilitation II: Proceedings of the 3rd International Conference on NeuroRehabilitation (ICNR2016), October 18-21, 2016, Segovia, Spain*, pages 153–157. Springer.
- [112] Cosima Prahm, Alexander Schulz, Benjamin Paaßen, Johannes Schoisswohl, Eugenijus Kaniusas, Georg Dorffner, Barbara Hammer, and Oskar Aszmann. Counteracting electrode shifts in upper-limb prosthesis control via transfer learning. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 27(5):956–962, 2019.
- [113] Christopher L Pulliam, Joris M Lambrecht, and Robert F Kirsch. Emg-based neural network control of transhumeral prostheses. *Journal of Rehabilitation Research and Development*, 48(6):739, 2011.
- [114] Haomeng Qiu, Yan Chen, Zhitao Chen, Chaojie Yang, Sihan Wu, Fanglin Li, and Longhan Xie. Feasibility of wrist-worn, cancelable, real-time biometric authentication via hd-semg and dynamic gestures. *IEEE Transactions on Instrumentation and Measurement*, 72:1–11, 2023.
- [115] Fayez Qureshi and Sridhar Krishnan. Wearable hardware design for the internet of medical things (iomt). *Sensors*, 18(11):3812, 2018.
- [116] Neta Rabin, Maayan Kahlon, Sarit Malayev, and Anat Ratnovsky. Classification of human hand movements based on emg signals using nonlinear dimensionality reduction and data fusion techniques. *Expert Systems with Applications*, 149:113281, 2020.
- [117] Sumit A Raurale, John McAllister, and Jesús Martínez Del Rincón. Emg biometric systems based on different wrist-hand movements. *IEEE Access*, 9:12256–12266, 2021.

- [118] Mamun Bin Ibne Reaz, M Sazzad Hussain, and Faisal Mohd-Yasin. Techniques of emg signal analysis: detection, processing, classification and applications. *Biological procedures online*, 8(1):11–35, 2006.
- [119] Arun Ross and Anil Jain. Information fusion in biometrics. *Pattern recognition letters*, 24(13):2115–2125, 2003.
- [120] Arun A Ross, Karthik Nandakumar, and Anil K Jain. *Handbook of multibiometrics*, volume 6. Springer Science Business Media, 2006.
- [121] Sherif Said, Abdullah S Karar, Taha Beyrouthy, Samer Alkork, and Amine Nait-ali. Biometrics verification modality using multi-channel semg wearable bracelet. *Applied Sciences*, 10(19):6960, 2020.
- [122] Walter Scheirer, Anderson Rocha, Ross Micheals, and Terrance Boulton. Robust fusion: Extreme value theory for recognition score normalization. In *European Conference on Computer Vision*, pages 481–495. Springer.
- [123] Erik Scheme and Kevin Englehart. Electromyogram pattern recognition for control of powered upper-limb prostheses: state of the art and challenges for clinical use. *Journal of Rehabilitation Research and Development*, 48(6), 2011.
- [124] Xinjun Sheng, Bo Lv, Weichao Guo, and Xiangyang Zhu. Common spatial-spectral analysis of emg signals for multiday and multiuser myoelectric interface. *Biomedical Signal Processing and Control*, 53:101572, 2019.
- [125] Siho Shin, Jaehyo Jung, and Youn Tae Kim. A study of an emg-based authentication algorithm using an artificial neural network. In *2017 IEEE SENSORS*, pages 1–3. IEEE.
- [126] Siho Shin, Mingu Kang, Jaehyo Jung, and Youn Tae Kim. Development of miniaturized wearable wristband type surface emg measurement system for biometric authentication. *Electronics*, 10(8):923, 2021.
- [127] Ryohei Shioji, Shin-ichi Ito, Momoyo Ito, and Minoru Fukumi. Personal authentication and hand motion recognition based on wrist emg analysis by a convolutional neural network. In *2018 IEEE International Conference on Internet of Things and Intelligence System (IOTAIS)*, pages 184–188. IEEE.
- [128] Edward Shwedyk, R Balasubramanian, and RN Scott. A nonstationary model for the electromyogram. *IEEE Transactions on Biomedical Engineering*, (5):417–424, 1977.

- [129] C Sinderby, L Lindstrom, and AE Grassino. Automatic assessment of electromyogram quality. *Journal of Applied Physiology*, 79(5):1803–1815, 1995.
- [130] Mirko Stanić. Continuous user verification based on behavioral biometrics using mouse dynamics. In *Proceedings of the ITI 2013 35th International Conference on Information Technology Interfaces*, pages 251–256. IEEE.
- [131] Toronto Star. "cyberattack on biometric data poses security risks at border, documents warn.", 2016.
- [132] Didier Staudenmann, Idsart Kingma, Dick F Stegeman, and Jaap H van Dieën. Towards optimal multi-channel emg electrode configurations in muscle force estimation: a high density emg study. *Journal of Electromyography and Kinesiology*, 15(1):1–11, 2005.
- [133] Abdulhamit Subasi. Classification of emg signals using combined features and soft computing techniques. *Applied soft computing*, 12(8):2188–2198, 2012.
- [134] Abdulhamit Subasi and Saeed Mian Qaisar. Surface emg signal classification using tqwt, bagging and boosting for hand movement recognition. *Journal of Ambient Intelligence and Humanized Computing*, pages 1–16, 2020.
- [135] Mahmoud Tavakoli, Carlo Benussi, Pedro Alhais Lopes, Luis Bica Osorio, and Anibal T de Almeida. Robust hand gesture recognition with a double channel surface emg wearable armband and svm classifier. *Biomedical Signal Processing and Control*, 46:121–130, 2018.
- [136] Dennis Tkach, He Huang, and Todd A Kuiken. Study of stability of time-domain features for electromyographic pattern recognition. *Journal of Neuroengineering and Rehabilitation*, 7(1):21, 2010.
- [137] Bahareh Tolooshams and Ning Jiang. Robustness of frequency division technique for online myoelectric pattern recognition against contraction level variation. *Frontiers in Bioengineering and Biotechnology*, 5:3, 2017.
- [138] Turker Tuncer, Sengul Dogan, and Abdulhamit Subasi. Surface emg signal classification using ternary pattern and discrete wavelet transform based feature extraction for hand movement recognition. *Biomedical Signal Processing and Control*, 58:101872, 2020.
- [139] Adilbek Turgunov, Kudratjon Zohirov, and Bobur Muhtorov. A new dataset for the detection of hand movements based on the semg signal. In *IEEE 14th International Conference on Application of Information and Communication Technologies (AICT)*, pages 1–4. IEEE, 2020.

- [140] Marina M-C Vidovic, Han-Jeong Hwang, Sebastian Amsüss, Janne M Hahne, Dario Farina, and Klaus-Robert Müller. Improving the robustness of myoelectric pattern recognition for upper limb prostheses by covariate shift adaptation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 24(9):961–970, 2015.
- [141] Alex Ming Hui Wong, Masahiro Furukawa, Hideyuki Ando, and Taro Maeda. Dynamic hand gesture authentication using electromyography (emg). In *2020 IEEE/SICE International Symposium on System Integration (SII)*, pages 300–304. IEEE.
- [142] Jian Wu, Lu Sun, and Roozbeh Jafari. A wearable system for recognizing american sign language in real-time using imu and surface emg sensors. *IEEE journal of biomedical and health informatics*, 20(5):1281–1290, 2016.
- [143] Qinghan Xiao. Technology review-biometrics-technology, application, challenge, and computational intelligence solutions. *IEEE Computational Intelligence Magazine*, 2(2):5–25, 2007.
- [144] Dezhen Xiong, Daohui Zhang, Xingang Zhao, and Yiwen Zhao. Deep learning for emg-based human-machine interaction: A review. *IEEE/CAA Journal of Automatica Sinica*, 8(3):512–533, 2021.
- [145] Hisaaki Yamaba, Tokiyoshi Kurogi, Kentaro Aburada, Shin-Ichiro Kubota, Tetsuro Katayama, Mirang Park, and Naonobu Okazaki. On applying support vector machines to a user authentication method using surface electromyogram signals. *Artificial Life and Robotics*, 23(1):87–93, 2018.
- [146] Jun Yang, Rong Yan, and Alexander G Hauptmann. Adapting svm classifiers to data with shifted distributions. In *Seventh IEEE international conference on data mining workshops (ICDMW 2007)*, pages 69–76. IEEE.
- [147] Wencheng Yang, Song Wang, Jiankun Hu, Guanglou Zheng, and Craig Valli. Security and accuracy of fingerprint-based biometrics: A review. *Symmetry*, 11(2):141, 2019.
- [148] Lin Yao, Jianjun Meng, Dingguo Zhang, Xinjun Sheng, and Xiangyang Zhu. Combining motor imagery with selective sensation toward a hybrid-modality bci. *IEEE Transactions on Biomedical Engineering*, 61(8):2304–2312, 2013.
- [149] Pan Yao, Kaifeng Wang, Weiwei Xia, Yusen Guo, Tiezhu Liu, Mengdi Han, Guangyang Gou, Chunxiu Liu, and Ning Xue. Effects of training and calibration data on surface electromyogram-based recognition for upper limb amputees. *Sensors*, 24(3):920, 2024.

- [150] Aaron J Young, Levi J Hargrove, and Todd A Kuiken. Improving myoelectric pattern recognition robustness to electrode shift by changing interelectrode distance and electrode configuration. *IEEE Transactions on Biomedical Engineering*, 59(3):645–652, 2011.
- [151] Xiangyang Zhu, Jianwei Liu, Dingguo Zhang, Xinjun Sheng, and Ning Jiang. Cascaded adaptation framework for fast calibration of myoelectric control. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 25(3):254–264, 2016.

Appendices

Appendix A

Table A.1: The unicode classification performance of each gesture for the Normal Test, Leaked Test and Identification scenarios.

UNI-CODE CLASSIFICATION PERFORMANCE			
Gesture	Normal Test (EER)	Leaked Test (EER)	Identification (FRR)
LP	0.017	0.040	0.182
TA	0.032	0.036	0.152
TLFO	0.035	0.052	0.217
TIFO	0.028	0.039	0.212
TLFE	0.026	0.037	0.081
TIFE	0.031	0.043	0.162
IMFE	0.040	0.046	0.192
LFE	0.031	0.051	0.172
IFE	0.032	0.039	0.248
TE	0.029	0.030	0.237
WF	0.025	0.056	0.126
WE	0.024	0.027	0.076
FS	0.028	0.028	0.157
FP	0.023	0.027	0.162
HO	0.021	0.023	0.076
HC	0.027	0.032	0.232

Appendix B

Table B.1: Within Day Identification Performance of Gestures

Sl No	Gesture	Rank-1 Accuracy
1	LP	83.98
2	TA	92.64
3	TLFO	93.07
4	TIFO	88.46
5	TLFE	93.22
6	TIFE	94.23
7	IMFE	91.05
8	LFE	91.34
9	IFE	91.63
10	TE	92.64
11	WF	93.07
12	WE	88.02
13	FS	90.62
14	FP	91.05
15	HO	91.77
16	HC	90.04

The individual gesture performance is utilized as weights for the majority voting decision level fusion.

Appendix C

Table C.1: GRABMyo Dataset Participant Demographics and Measurements

Sl	ID	Age	Sex	Forearm Length	Forearm circumference	Wrist circumference
1	participant1	27	F	26.5	26.0	16.5
2	participant2	25	M	26.0	26.0	17.5
3	participant3	24	F	25.0	22.0	15.0
4	participant4	29	F	25.0	21.6	16.2
5	participant5	26	M	27.0	30.0	18.6
6	participant6	26	F	23.0	21.0	13.9
7	participant7	25	M	26.5	25.4	17.4
8	participant8	27	F	26.5	24.0	17.0
9	participant9	25	M	26.0	24.5	17.0
10	participant10	32	F	23.5	23.0	16.0
11	participant11	29	F	23.0	25.0	17.5
12	participant12	29	F	23.0	19.0	14.0
13	participant13	25	F	24.0	23.0	16.8
14	participant14	27	M	27.2	25.0	17.2
15	participant15	26	F	23.0	22.0	16.0
16	participant16	28	M	25.2	25.0	16.0
17	participant17	25	M	23.4	25.0	16.2
18	participant18	31	M	24.5	25.2	16.2
19	participant19	24	F	24.0	21.0	16.0
20	participant20	19	F	24.2	26.7	18.0
21	participant21	24	M	24.5	26.5	16.5
22	participant22	24	M	25.0	23.2	15.4

23	participant23	23	F	23.5	22.8	16.2
24	participant24	26	M	24.0	23.0	15.5
25	participant25	22	F	24.0	24.0	16.5
26	participant26	24	M	28.0	26.0	16.0
27	participant27	28	M	28.4	27.0	16.5
28	participant28	28	M	25.3	25.0	16.0
29	participant29	32	F	24.0	20.2	14.0
30	participant30	30	M	26.0	24.0	16.2
31	participant31	27	M	28.4	28.9	19.4
32	participant32	25	M	27.4	23.1	16.0
33	participant33	27	M	24.5	24.0	15.5
34	participant34	23	F	25.0	26.0	17.2
35	participant35	24	M	27.5	25.2	16.5
36	participant36	27	M	28.0	24.0	15.6
37	participant37	27	M	26.0	21.0	14.0
38	participant38	28	F	22.8	23.2	15.0
39	participant39	24	F	21.8	20.0	14.0
40	participant40	28	F	23.0	26.0	18.0
41	participant41	34	M	27.5	25.0	16.0
42	participant42	24	F	24.5	23.0	15.5
43	participant43	27	M	26.0	25.0	15.5
Average		26.395		25.153	24.105	16.186
SD		2.871		1.745	2.278	1.217

Appendix D

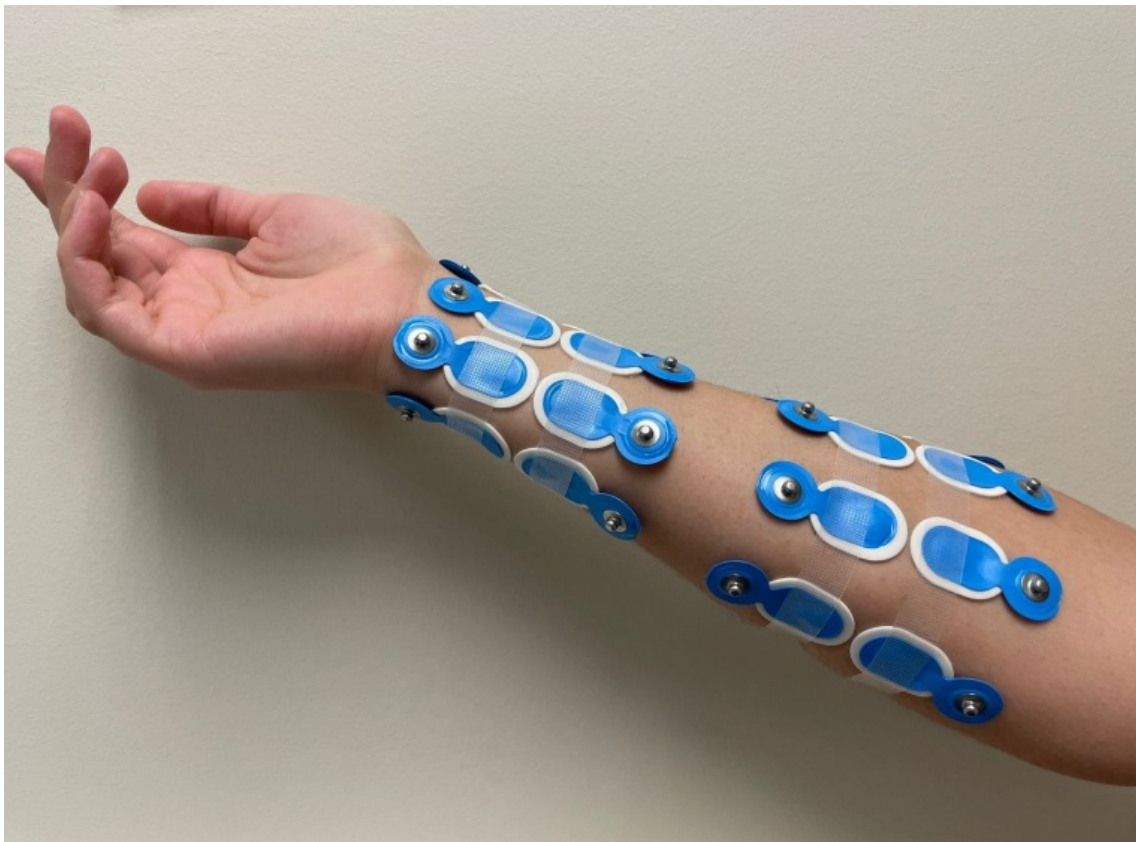


Figure D.1: The electrode locations on the wrist (12) and forearm (16) as used in the GRABMyo Dataset