



Department of Computer Science
Memorial University of Newfoundland & Labrador
240 Prince Philip Dr, St. John's, Newfoundland and Labrador, Canada

Machine Learning and Processing Techniques for the Enhancement of Hand Gesture Recognition of Forcemyography and Electromyography Signals

Authored by
© **Mohammed Asfour**

Supervised by
Dr. Xianta Jiang

A thesis submitted to the School of Graduate Studies
to partially fulfill the requirements for the Masters of
Science degree.

December 2022

Abstract

Hand gesture recognition is the primary driver of many applications across user groups. It uses machine learning classifiers to classify users' data, most prominently Electromyography (EMG) or Forcemyography (FMG). Whereas EMG sensors detect signals going down the arm to the muscles, FMG sensors measure the pressure change on the arm's skin. Nevertheless, many inconsistencies impact gesture recognition drastically. For instance, gesture recognition for the same user, intra-subject, is affected by the duration between the collected signals for classifiers' training and the classified signals, requiring more data from the user. A more significant hurdle is inter-subject gesture error, in which classifiers are trained on signals from one or more subjects perform exceptionally poorly on the signals of another. These issues arise due to the uniqueness of such signals per person and their variance through time. We offer methods to encounter several downsides of EMG and FMG. We propose a machine learning pipeline that yields features of consistent performance across various classifier types and reduces intra-subject signal variance. To tackle other intra-subject errors, we offer a ranking for what we define as the feature-classifier compatibility relationship that controls the recognition performance. The methods are tested on FMG and EMG, respectively, and enhanced gesture recognition.

General Summary

Recognizing hand gestures is essential to many applications, from virtual reality to artificial limbs. These applications use sensors to detect signals transmitted to the hand muscles, Electromyography (EMG), or measure deflection on the arm skin, Force myography (FMG), while performing these gestures. However, many challenges arise in these applications. Due to the change of the measured signals for a single person with time, recognition performance can also change with time, requiring more samples from the application user. A bigger problem is the uniqueness of such signals for each person, prohibiting using other users' signals to boost the gesture recognition of the end-users of these applications. We introduce machine learning and signal processing methods to limit these factors for EMG and FMG. We subsequently use a series of machine learning to process FMG signals, making their classification more consistent. As for EMG, we introduce a compatibility relationship that controls recognition performance.

Co-Authorship Statement

The research in this thesis is the product of the collaborative efforts of Mohammed Asfour, Dr. Xianta Jiang, and Dr. Carlo Menon. Mohammed Asfour is the primary investigator of the research in this thesis. Dr. Jiang and Dr. Menon are senior researchers supervising and directing the research.

The authors' roles in the research, whose results are presented, are as follows:

- **Conceptualization:** Mohammed Asfour, and Xianta Jiang
- **Data curation:** Xianta Jiang, and Carlo Menon
- **Formal analysis:** Mohammed Asfour
- **Funding acquisition:** Xianta Jiang
- **Investigation:** Mohammed Asfour
- **Methodology:** Mohammed Asfour
- **Supervision:** Xianta Jiang
- **Writing:** Mohammed Asfour
- **Review and editing:** Mohammed Asfour, and Xianta Jiang

Mohammed Asfour has obtained the consent of the two senior researchers, the co-authors, to publish the research and its findings as part of his thesis.

Acknowledgements

The research conducted as a part of this thesis was funded by The Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery grant under Grant RGPIN-2020-05525.

Topics and findings of this thesis chapters have been published previously in peer-reviewed journals in other versions, and the publishers' approval was granted. Chapter 2 has been published as "Asfour, M.; Menon, C.; Jiang, X. A Machine Learning Processing Pipeline for Reliable Hand Gesture Classification of FMG Signals with Stochastic Variance. *Sensors* 2021, 21, 1504. <https://doi.org/10.3390/s21041504>" [1].

The third chapter has been published under the title "Asfour, M.; Menon, C.; Jiang, X. Feature–Classifier Pairing Compatibility for sEMG Signals in Hand Gesture Recognition under Joint Effects of Processing Procedures. *Bioengineering* 2022, 9, 634. <https://doi.org/10.3390/bioengineering9110634>" [2].

The ethics committee of Simon Fraser University and the data custodian approved us to dissect their previously collected data in this research. We also obtained the Interdisciplinary Committee on Ethics in Human Research (ICEHR) approval from the Memorial University of Newfoundland for the secondary use of data under application number 20230479-SC.

This thesis would not have been possible without the supervision and ongoing advice of Dr. Xianta Jiang for the entirety of my program, on and off the research. I would also like to thank my lab mates, who gave me feedback on my experiments.

I thank Allah for giving me this opportunity and blessing me with my family, especially my parents, to whom I dedicate this achievement for their lifelong support. Lastly, I thank my friends, who were essential in this journey.

Contents

Abstract	ii
General Summary	iii
Co-Authorship Statement	iv
Acknowledgements	v
List of Figures	viii
List of Tables	x
1 Overview	1
1.1 Background	1
1.2 Problem Statement	2
1.3 Literature Review	3
1.4 Thesis Work	5
1.5 Contributions	7
2 Machine Learning Processing of FMG Signals for Reduced Intra-subject Gesture Recognition Errors	9
2.1 Introduction	9
2.2 Materials and Methods	11
2.2.1 Forcemyography Data	11
2.2.2 Design of the Machine Learning Pipeline	12

2.2.3	Classification Models Used for Evaluation	15
2.2.4	Statistical and Validation Methods	17
2.3	Results	17
2.4	Discussion	23
2.4.1	Gesture Recognition Enhancement	23
2.4.2	Similarity of Processed Distributions	25
2.4.3	Limitation and Future Work	26
2.5	Conclusion	27
2.6	Co-Authorship Statement	27
3	Compatibility of Feature-Classifier Pairs for EMG Hand Gesture Recognition under Joint Processing Procedures	29
3.1	Introduction	29
3.2	Materials and Methods	31
3.2.1	Electromyography Data	31
3.2.2	Hypotheses Testing and Conducting Experiments	33
3.2.3	Experiments' Control Settings for Generalizable Results	38
3.3	Results	41
3.3.1	EMG Signal Normalization	41
3.3.2	Grid Search for tunable Features and Classifiers	41
3.3.3	Processing Window Size Impact on Recognition	42
3.3.4	Feature-Classifier Compatibility and Accuracy Ranking	43
3.4	Discussion	45
3.5	Conclusion	50
3.6	Co-Authorship Statement	50
4	Conclusion	52
	Bibliography	56
A	Interdisciplinary Committee on Ethics in Human Research (ICEHR) Approval	67

List of Figures

2.1	Diagram of the proposed machine learning pipeline.	12
2.2	Connections created by UMAP on participant two training data prior to and post UMAP optimization step with enlargement on gesture G7 neighborhood.	14
2.3	Feature spaces after applying each pipeline step for all participant 2 data gestures, displaying the first two dimensions for each feature space.	18
2.4	Feature spaces after applying each pipeline step, highlighting medium wrap gesture (G3) of participant 2 training data, showing the first two dimensions for each feature space.	19
2.5	Classifiers' accuracy averaged over all participants for each test session using raw and pipeline features. Error bars resemble one standard deviation.	20
2.6	Confusion matrices for QDA, FC-NN, and KNN displaying per-class accuracy on the pipeline's features for participant two's test sessions.	21
2.7	Comparison of QDA accuracy using raw and pipeline features for each test session of all participants.	22
3.1	The MyoSystem 1400L EMG acquisition device.	32
3.2	EMG samples acquired from participant 4.	33
3.3	A diagram of the categories in the dataset.	33
3.4	Accuracy results of Logistic Regression (LR), Linear Discriminant Analysis (LDA), and Quadratic Discriminant Analysis (QDA) averaged on participants 4, 5, and 6 data using no-tuning features and varying window size.	41

3.5	Accuracy results of Logistic Regression (LR), Linear Discriminant Analysis (LDA), and Quadratic Discriminant Analysis (QDA) with multiple WA thresholds on varying window sizes averaged on EMG data of participants 4, 5, and 6.	42
3.6	Accuracy results of Logistic Regression (LR), Linear Discriminant Analysis (LDA), and Quadratic Discriminant Analysis (QDA) with multiple SSC thresholds on varying window sizes averaged on EMG data of participants 4, 5, and 6.	43
3.7	Features test recognition accuracy for different window sizes of all classifiers averaged on sEMG data of participants 4, 5, and 6.	44
3.8	Distribution of accuracy results for all feature-model pairs for grasp label group averaged on all participants' data from the first session.	45
3.9	Feature accuracy ranking on the 1250 ms window with classifiers' sub-ranking averaged on all participants' grasp data.	46
3.10	Classifier accuracy ranking on the 1250 ms window with features' sub-ranking averaged on all participants' grasp data.	46
3.11	Accuracy distribution per feature on Grasp Gestures, Sign Language Gestures, and Unique Movements Gestures averaged on all classifiers and all participants.	47

List of Tables

2.1	Tuned UMAP hyperparameters.	15
2.2	Tuned Neural Network hyperparameters.	16
2.3	Mean accuracy for the pipeline features and features of FDA, PCA, and UMAP individually as standalone preprocessing methods. The accuracy results are averaged over all participants and rounded to 1 decimal place.	23
3.1	Values used in grid search for classifiers hyperparameters.	40
3.2	Classifiers' optimal hyperparameters after performing grid search using sEMG data of participants 4, 5, and 6.	43

Chapter 1

Overview

1.1 Background

Various fields depend on hand gesture recognition to deliver realistic user experiences, from entertainment purposes such as virtual reality (VR) [3] to critical applications such as remote control of robotic arms and upper-limb prosthetics [4] [5] [6]. Gesture recognition can be based on several data types, such as visual data from cameras [7], motion data obtained from worn-on gyroscopes [8], or muscle activation data such as surface Electromyography (EMG) [9] and Forcemyography (FMG) [10]. The data is used to train machine learning classifiers to predict the gestures of future unseen data.

Electromyography (EMG) is the collection of signals transmitted down the forearm to the hand muscles. It is performed using electrodes worn on the user's forearm. On the other hand, Forcemyography (FMG) sensors measure the pressure change on the forearm skin due to muscles' contraction and relaxation. Compared to the other methods, EMG and FMG do not suffer from vital problems such as occlusion, and they are more sensitive to subtle differences in hand gestures and power exerted [11] [12]. At the same time, EMG and FMG are non-invasive methods compared to more complex biosignals acquisition methods.

1.2 Problem Statement

Electromyography and forcemyography are the main methods used in critical applications, such as upper-limb prosthetics. Nevertheless, it is seen that prosthetics users prefer to use artificial mechanical limbs over electrically-powered ones [13], more than 70% of users in some studies [14]. Their disfavor of electrical prostheses is due to the weak reproducibility of correct gesture recognition arising from hindrances of such methods [15] [16]. For instance, electromyography is susceptible to high noise levels because of sweat on the skin [17]. Another downside is the high cost compared to the rest of the methods. Even though FMG does not suffer from these problems, being relatively cheaper and more robust to sweat, it suffers from other factors. FMG is prone to signal changes over time due to muscle fatigue [18] or sensors displacement. These limiting factors, and others, contribute to different gesture recognition errors using EMG or FMG.

Intra-subject errors are significantly prominent in EMG and FMG gesture recognition, causing recognition performance to decline for the same user, especially with time [18] [19] [11]. The drop can be attributed to the change of muscle activation patterns for the same gesture due to fatigue, among other sources, which makes gesture recognition inconsistent. For instance, the inter-session error is an intra-subject error that occurs when classifiers' training data is from one session and the test or prediction data is from another.

A more complex issue is the inter-subject gesture recognition error, referring to the plunge of recognition accuracy when machine learning classifiers are trained on data from one or more users and tested on another [20]. This inter-subject error can be tracked to the uniqueness of EMG and FMG signals per user, thus prohibiting any benefits from incorporating more data from other users.

Such errors question whether it is possible to reduce them and provide a more reliable recognition experience. Because machine learning methods have enabled breakthroughs in different fields, they are the main research procedures in EMG and FMG processing literature.

1.3 Literature Review

Several studies were conducted on different EMG and FMG gesture recognition errors, most notably inter-subject and intra-subject errors, to analyze and minimize them. These studies aimed to enhance user-specific classifiers of hand gestures and mainly reduce intra-user errors for EMG.

An example of such errors can be viewed in a study by Pale et al. [15], who investigated the similarity of muscle synergies extracted from EMG signals from different sessions of the Nipapro dataset 6. Muscle synergy is a hypothesized concept regarding how the nervous system controls the muscles simultaneously and can be considered another representation of the EMG signals. Their empirical investigation concluded that the change of the muscle synergies extracted from EMG varies with sessions, which could not be explained by sensor repositioning or other factors. Their results supported their conclusion as they reported inter-session synergies to have 0.2–15% variance accounted for (VAF). In contrast, intra-session synergies had 0.1–2.5% VAF, highlighting a much higher inter-session synergies variance than the intra-session synergies’.

Despite Forcemyography being seen as a possible alternative to electromyography, it suffers from inconsistent, user-specific errors similar to EMG, such as the inter-session errors mentioned.

For instance, Belyea et al. [21] have compared gesture recognition performance based on FMG and EMG signals. In their comparison, they used evaluation metrics such as throughput, a performance index from Fitts’ law which is a model of human movement used in human-computer interaction. They also defined a path efficiency metric to measure the quality of the hand’s path to achieve the required gesture. For classification tasks, they reported an average of 90.02% for FMG versus 75.1% for EMG on the throughput metric and 87.2% for FMG compared to 83.2% for EMG on their defined path efficiency metric.

In another study, Jiang et al. [22] got similar findings while comparing electromyography-based and forcemyography-based gesture recognition, as mentioned in Section 2.1. Their results revealed that FMG signals are more indicative of the hand gesture in the case of linear discriminant analysis (LDA) classification than EMG. However, the LDA classifier also ex-

perienced a drop in recognition accuracy on FMG as on EMG signals. Using eight sensors for each of the two methods, they reported 91.2% and 83.5% recognition accuracy from their first test session and a later test session for FMG. In contrast, they obtained 84.6% and 79.1% accuracy for EMG from the same test sessions mentioned.

Both studies pointed out the same inter-session error in FMG and EMG. In addition to this error, other user-specific errors are evident in FMG signal gesture recognition.

An example of user-specific errors is shown in a study by Jiang et al. [18], in which they investigated hand gesture recognition using forcemyography across different levels of force exerted while performing the gestures. They used one data collection session to train an LDA classifier per subject and tested it on two separate testing data collection sessions. Besides a 10% drop in recognition accuracy between the two sessions, they found that the minimum force exerted yielded 70% accuracy, which rose to about 86% with increased force, then decreased for stronger force levels. This study showcased another user-specific error of FMG signals that arises from varying force levels used to apply the exact gesture for the same person.

Intra-subject errors aside, inter-subject gesture recognition offers major benefits if applied successfully. However, it reduces inter-subject gesture classification to near randomness if not used adequately. Inter-subject gesture recognition studies are primarily concerned with EMG signals and oriented toward domain adaptation techniques known in machine learning literature as the compensation of the shift between two data distributions. As for gesture recognition, domain adaptation uses machine learning to adapt classifiers trained on specific users' data to the target user data.

One study proposed a domain shift technique by Du et al. [23] used a neural network to adapt classifiers trained on EMG data collected from several subjects to EMG data collected from a target subject. They obtained an inter-subject gesture classification accuracy as high as 55.3% on the CapgMyo-b electromyography dataset. However, they used half of the test data of the target subject to perform the domain adaptation and evaluated the classifiers using the other half. As a strong counterpoint, the data used for domain adaptation could have been used to train a user-specific classifier that maintained or even surpassed the domain adaptation performance.

In addition to the counterpoint stated, Marano et al. [24] investigated several proposed domain adaptation techniques for inter-subject hand gesture classification using electromyography. They found that the reported results could only be achieved in specific experimental settings and are heavily influenced by excluded factors, such as classifiers' hyperparameters tuning. They also concluded that the amount of data used for domain adaptation in these techniques could train user-specific classifiers that surpassed domain adaptation methods in most scenarios.

As seen from the studies discussed, there are two main reasons, among others, that impede the adoption of EMG and FMG hand gesture recognition applications. The first arises from the inconsistency of classification performance due to intra-subject variance of the signals, as proved by the drop in accuracy between sessions, different force levels, and other factors [18]. Thus, the survey of prosthetics users, for example, preferred artificial mechanical limbs to EMG ones due to the abovementioned inconsistency [14]. The latter comes down to the poor generalization of the results of the controlled experiment into other conditions. This is seen in the performance downgrade of the classifiers in real-time applications [25] or the negation of the classifier's role when evaluating the quality of the signal processing methods.

1.4 Thesis Work

This thesis aims to limit the intra-subject errors of FMG and EMG signals to boost the reproducibility of gesture recognition performance and make it more consistent. Reducing these errors enriches the user experience of gesture-recognition applications, such as making electrically-powered prosthetics more appealing to mechanical prosthetics users or enhancing virtual reality experiences.

The thesis proposes two methods to address the mentioned errors, mainly intra-subject errors. We developed a machine learning pipeline to process signals into robust signal representation, reducing the user's classification error across different classifiers. The pipeline increased the forcemyography signals' reliability and reduced recognition errors per user. The second method ranks the performance of feature-classifier pairs, showing that compatibility

between a feature and a classifier vastly influences the recognition accuracy despite the lack of investigation of that compatibility in the literature. We show that the suggested compatible pairings acquire the highest recognition accuracy for EMG signals alongside normalization and preprocessing steps that reduce intra-subject recognition error further.

Chapter 2 proposes a pipeline of machine learning methods that process force myography signals to gain preferred characteristics, boosting gesture classifiers. These characteristics enable the pipeline to reduce the stochastic variance in FMG signals for enhanced gesture recognition. The pipeline consists of Fisher's discriminant analysis (FDA) which projects the data to separate signals of different gestures, making them easier to classify correctly. FDA is then followed by principal component analysis (PCA), aiming to remove any correlation between the new dimensions. Correlation removal is intended to raise the performance of less complex classifiers and linear models. The inter-dependency between dimensions requires a more complex decision boundary to separate signal data according to their correct gestures. Finally, we apply Uniform Manifold Approximation and Projection (UMAP), one of many machine learning techniques known as manifold learning methods. These methods hypothesize that data in $n - dimensional$ space lie on a manifold on which the data distribution is simpler, in the sense that data of different labels are more separated. This last step helps reduce the complexity of the distribution and finds a better non-linear transformation of the space.

The proposed machine learning pipeline achieves desirable attributes for the FMG data as all classifiers' accuracy was boosted to reach the same level. The classifiers' similar performances, regardless of their complexity, proved the processed signals to be more linearly separable in the data space according to their gestures and are subject to the same amount of error. This pipeline enables a more consistent gesture recognition of FMG data regardless of the classifier, benefiting low-computation devices, and yielding better consistency through time. Therefore, the pipeline reduced intra-subject errors within each test session and minimized the intra-session error.

Chapter 3 discusses the reliability of the classifiers and signal filters chosen to process and classify electromyography signals. Due to certain limitations of the EMG gesture recognition literature studies, we investigate the pairing choice of EMG features and classification methods.

Most studies use a single feature to test the performance of several classifiers or vice versa.

We find a compatibility relationship between the feature used and the selected classifier that primarily determines the gesture recognition performance for EMG signals. The compatibility shows that the pairing choice of the feature-classifier pair can influence results more than the feature type used or the complexity of the classifier. All the pairs of ten classifiers and 12 features are investigated and ranked. This ranking provides guidelines for future research and applications of EMG gesture recognition, showing which pairs outperform the others. Most importantly, the ranking proposes that linear classification models can outperform ensemble models when paired with the proper features. Thus, it allows low-computational devices to use linear models without sacrificing recognition accuracy. The chapter also investigates the best preprocessing steps over multiple scenarios. The recognition performance change across four normalization ranges is included, along with the recognition performance of 2 features during their tuning. In addition, the performance difference with varying filtering window sizes and the optimal models' hyperparameters are added. The optimal window size is much above the limit for real-time gesture recognition, 300ms. However, smaller window sizes can achieve real-time recognition with a slight decrease in performance. Using the suggested compatible feature-classifier pairs gives the most accurate recognition compared to other pairs. In addition, the proposed signal normalization and window size reduce the intra-subject recognition errors for each user even further.

1.5 Contributions

The work of this thesis has been published partially in other versions in peer review journals. We have the approval of the publishers to use the published findings in the thesis. Published versions of this research chapters:

- Chapter 2 proposes a signal processing pipeline employing a manifold learning method to produce a robust signal representation to boost hand gesture classifiers' performance and reduce intra-subject errors. This chapter is published as **Asfour, M.; Menon, C.; Jiang, X. A Machine Learning Processing Pipeline for Reliable Hand Gesture Clas-**

sification of FMG Signals with Stochastic Variance. *Sensors* 2021, 21, 1504.

<https://doi.org/10.3390/s21041504>. [1]

- Chapter 3 explores the feature–classifier pairing compatibility for sEMG, showing that it is an unexplored primary determinant of gesture recognition accuracy. The proposed pairing ranking provides a guideline for choosing the proper feature or classifier in future research. This chapter is published as **Asfour, M.; Menon, C.; Jiang, X. Feature–Classifier Pairing Compatibility for sEMG Signals in Hand Gesture Recognition under Joint Effects of Processing Procedures. *Bioengineering* 2022, 9, 634. <https://doi.org/10.3390/bioengineering9110634> [2].**

Chapter 2

Machine Learning Processing of FMG

Signals for Reduced Intra-subject Gesture

Recognition Errors

2.1 Introduction

Many daily activities rely on hand gesture recognition, such as virtual reality (VR) [3], human-robot interaction [4,5] and prosthesis control [6]. Gesture recognition is performed by machine learning classifiers trained on pre-collected data from the users. Such data can be visual data from cameras [7], inertial data obtained from a gyroscope or accelerometer [8], or muscle activity data such as surface Electromyography (sEMG or EMG) [9,26]. Unlike most methods, muscle activity-based hand gesture recognition does not have occlusion problems. It can also detect fine motor hand gestures, and possibly gesture-performing strength [11,12].

Hand gesture recognition relies primarily on EMG [27,28] whose sensors are mounted on the upper limb to detect muscles' signals sent down the arm. EMG signal processing and classification have been investigated thoroughly for various applications [6,19,29–31]. Force myography (FMG) [32] is seen as an alternative with recent investigations in hand gesture recognition literature [10,33]. FMG-based hand gesture recognition uses force-resisting sensors on the forearm to capture the volumetric muscle changes while performing gestures [34].

Unlike EMG, FMG is robust to external electrical interference and sweating, inexpensive, and easy to use. Using sophisticated machine learning techniques, FMG can achieve competitive performance to EMG [35]. Jiang et al. [22] analyzed FMG sensors' performance in hand gesture classification compared to EMG. The results showed that using only 8 FMG sensors obtained gesture recognition accuracy on par with commercially available EMG prostheses in a controlled study.

However, EMG and FMG are prone to stochastic variation in their signals for the same gesture with time passage, leading to low inter-session classification performance [15, 16]. Feature engineering has been used to address this limitation previously [36, 37]. Tkach et al. [36] investigated stochastic variance of EMG signals and their effect on classification performance. They concluded that although feature engineering could combat the effect, it could not eliminate the error. They recommend that further research be conducted to improve the robustness of EMG signals. Ketykó et al. [37] inspected the variation of EMG signals between sessions and with multiple participants, also known as domain shift, which affects the accuracy negatively. They proposed a recurrent neural network (RNN) to process EMG. They analyzed their model's performance on public EMG datasets to enhance the recognition performance. However, using half the collected data trials, 50% of CapgMyo dataset [38], without limitations on their order in training, could have reduced the variance by itself. As described in their study, trials 1, 3, 5, 7, and 9 were chosen to train the classifier, reducing the variance effect compared to choosing trials 1, 2, 3, 4, and 5 for training.

This chapter introduces a novel pre-processing pipeline to reduce the stochastic variance of FMG signals in hand gesture classification, a type of intra-subject error. Firstly, Fisher's Discriminant Analysis (FDA) [39] clusters data points of the same gesture together while separating them from other gestures' data. Secondly, Principal Component Analysis (PCA) [40] eliminates correlation between data features. Lastly, Uniform Mapping and Approximation Projection (UMAP) [41] learns an internal data pattern that transforms them into enhanced features. The proposed method's performance and the robustness of the processed features were evaluated with five classifiers compared to raw FMG data.

2.2 Materials and Methods

2.2.1 Force myography Data

The data in this chapter was obtained from a previous study [18] investigating the force exertion effect on hand gesture classification. The Interdisciplinary Committee on Ethics in Human Research (ICEHR) at the Memorial University of Newfoundland approved the secondary use of the data for our research under application number 20230479-SC. The data was collected using an array of 16 force-sensitive resistors (FSR) in a custom-designed wristband. The data collection process was divided into a training session and two testing sessions, with 16 grasp gestures performed per session. The grasps were large diameter heavy wrap (G1), small diameter heavy wrap (G2), medium wrap (G3), adducted thumb wrap (G4), light tool (G5), thumb + 4 fingers pinch (G6), thumb + 3 fingers pinch (G7), thumb + 2 fingers pinch (G8), thumb + 1 finger pinch (G9), disk power grasp (G10), sphere power grasp (G11), disk precision grasp (G12), sphere precision grasp (G13), tripod grasp (G14), push (G15), and lateral pinch (G16).

In that study, signals from varying force levels were recorded from nine participants, seven males and two females, with a median age of (27 ± 6) years. All were right-handed and self-reported to be 100% functional with their working hands. All participants read and signed the consent form before the study, approved by Simon Fraser University to collect their data throughout 180 repetitions, four for training and eight for each of the two testing sessions. Data from the 16 FSR channels were obtained with a sampling rate of 15 Hz [18].

The training session's gestures were performed using natural grasping force, similar to naturally grasping an object in daily activities. Each repetition lasted 3 s, yielding 45 samples with a 15 Hz sampling rate. Thus the 16 gestures, with four repetitions each, yielded 2880 training samples for each participant.

Alternatively, the testing sessions' gestures were performed using eight grasping force levels purposely, yielding 5760 test samples for each testing session per participant.

A monitor displayed the object's image with the required exerted force level as a reference to guide the participants to match the force level on the monitor. There was only a short break between the sessions, and the band was not removed in between. This data collection protocol

was used for 9 participants to collect 16 gestures' data. For a more detailed description of the data collection protocol, please refer to [18].

2.2.2 Design of the Machine Learning Pipeline

The pipeline successfully employs FDA, PCA, and UMAP algorithms, as shown in Figure 2.1.

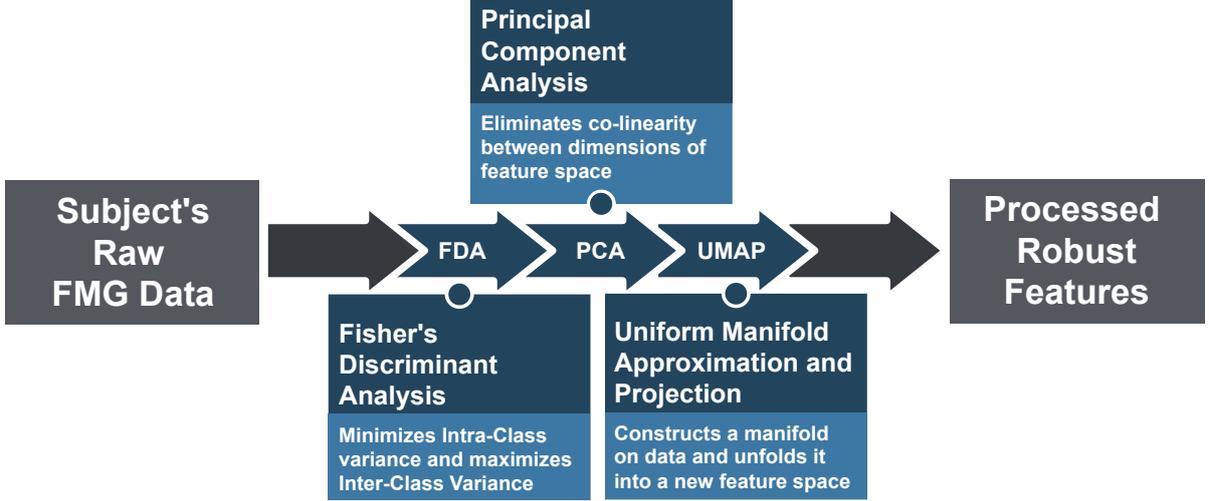


Figure 2.1: Diagram of the proposed machine learning pipeline.

- **FDA Model:** The raw FMG data firstly underwent Fisher's Discriminant Analysis (FDA) [39]. FDA finds the best linear projection that maximizes the separation of different classes' data, reducing the overlap between different classes. Maximizing the distances between different classes' points increases inter-class variance. In contrast, minimizing distances between the same class points decreases intra-class variance. FDA solves a constrained optimization function using Lagrangian multipliers as follows

$$L(\lambda, W) = \text{trace}(W^T S_B W) - \lambda (\text{trace}(W^T S_W W) - 1) \quad (2.1)$$

where S_B is the variance between classes, inter-class variance, and S_W is the variance within each of the classes, intra-class variance. Both of the variance matrices' dimensions are $d * d$, where d is the number of the original data dimensions. The solution W , containing the basis vectors of the transformation space, is given by

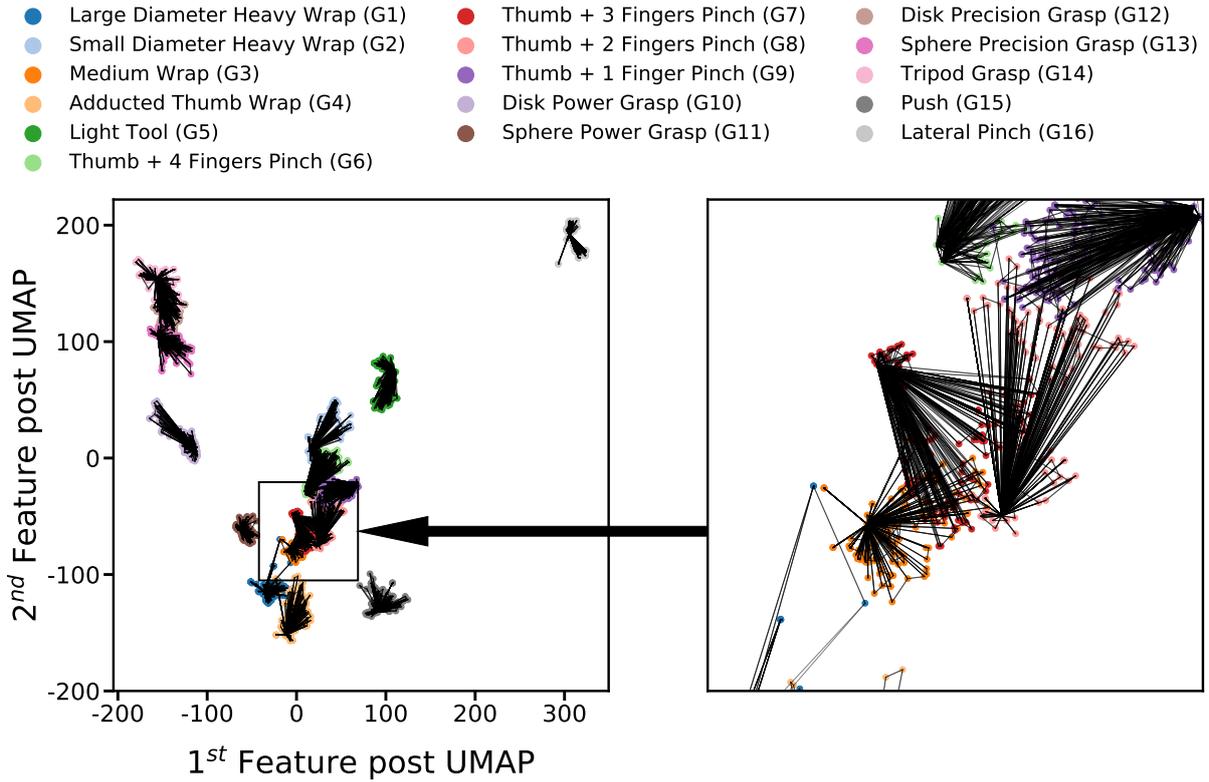
$$W = \text{eigen}(S_W^{-1} S_B) \quad (2.2)$$

- **PCA Model:** After FDA processing, PCA [40] was applied to remove the correlation between data dimensions. Removal of correlation significantly benefited the final step of the pipeline, UMAP. UMAP learned the underlying structure more efficiently without considering any inter-dimensional relationship. The principal components can be obtained from

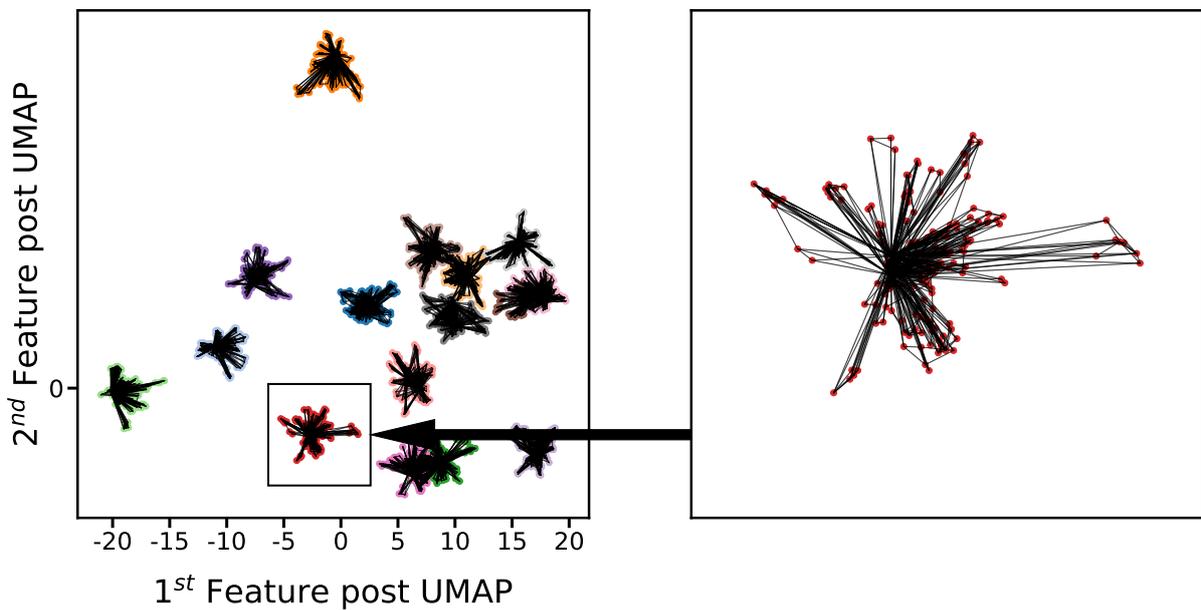
$$U = \text{eigen}\left(\frac{1}{n} (X - \mu)^T (X - \mu)\right) \quad (2.3)$$

where X_{n*d} is the matrix of n data samples in d -dimensional space, and μ_{1*d} is the mean of the data.

- **UMAP Model:** Finally, UMAP [41] was used to produce a more robust set of data features. This method assumes the data is distributed on a connected manifold. A manifold is a nonlinear surface that resembles a new Euclidean space if unfolded. UMAP connects data points to construct that manifold, followed by an optimization step to unfold the manifold to a simpler distribution. Its optimization is weighted between conserving the relative positions of points through the established connections and separating different classes after unfolding via the points' classes. The graph resulting from UMAP connections between training points for participant two after applying FDA and PCA is depicted in Figure 2.2a, whereas the same points distribution with the same connections after applying UMAP is shown in Figure 2.2b. UMAP [42,43] algorithm has many hyperparameters to tune its manifold and optimization. We prioritize finding the separation between different classes to reduce the stochastic variance of FMG signals. UMAP hyperparameters were tuned on participant two's data and are shown in Table 2.1.



(a) UMAP connections prior to the optimization step.



(b) UMAP connections post the optimization step.

Figure 2.2: Connections created by UMAP on participant two training data prior to and post UMAP optimization step with enlargement on gesture G7 neighborhood.

FDA is a linear transformation based on gesture classes, whereas PCA is a linear class-agnostic transformation. Both can be considered pre-processing steps for UMAP, the vital step

Table 2.1: Tuned UMAP hyperparameters.

UMAP Hyperparameter	Tuned Value
Number of Neighbors	2
Metric	Cosine Distance
Output Metric	Euclidean
Target Metric	Euclidean
Target Weight	0.75
Repulsion Weight	3.0
Embedding Initialization	Random
Minimum Distance	0.25

in the pipeline, giving consistent features throughout time sessions. Nevertheless, its performance is boosted by pre-eliminating specific characteristics using FDA and PCA.

Even though UMAP separates the classes, FDA linear transformation has made UMAP’s optimization easier and enhanced the set of possible nonlinear UMAP manifolds. By reducing the overlap using FDA, UMAP manifolds can give better results, which we discuss in detail. PCA is used in the pipeline for correlation elimination. We found that UMAP’s chosen manifold, given a non-correlated set of features, separates the classes more consistently throughout time. PCA has been coupled with UMAP [42, 43] in the literature for different purposes, such as data analysis and visualization, rather than the proposed pipeline.

2.2.3 Classification Models Used for Evaluation

Five machine learning classifiers are used to evaluate the performance after applying the proposed pipeline, compared to the raw FMG features [18, 22]. The models used in this chapter are listed below:

- **Linear Discriminant Analysis (LDA)** [44] has been widely used in gesture recognition [32, 45–47]. LDA assumes that each class’ data is normally distributed in the feature space with the same variance for all classes.
- **Quadratic Discriminant Analysis (QDA)** [48] is similar to LDA, assuming classes are normally distributed but with separate variances, giving a quadratic decision boundary.
- **Support Vector Machine with Radial Basis Kernel (SVM-RBF)** [49] uses a kernel to

transform data into another feature space before finding a linear decision boundary in it. The radial basis function transforms data into infinite-dimensional space, theoretically. The linear boundary in the transformed space is nonlinear in the original feature space.

- **Fully-Connected Neural Network (FC-NN)** [50] is the most complex of the used models. Only a few fully-connected layers were sufficient for this study. Neural networks have numerous hyperparameters to tune, making them highly flexible during design. The hyperparameters are tuned once to evaluate the pipeline subjectively in Table 2.2. Regularization techniques are used for a better generalization of test data.
- **K-nearest Neighbors (KNN)** [51] is one of the most basic classifiers, which uses neighboring labeled data points to classify unlabelled data.

Table 2.2: Tuned Neural Network hyperparameters.

Hyperparameter	Used Value
Learning Rate	0.001
Epochs	30
Batch Size	1024
Validation Split	0.2

The decision boundaries of these models have different orders of non-linearity. For instance, LDA has linear boundaries, and FC-NN has highly nonlinear boundaries. In contrast, KNN uses a distance metric instead of decision boundaries. The higher the classifier’s non-linearity or flexibility, the higher its capacity or complexity is said to be. Enhancement in all classifiers’ performance would suggest that the pipeline yields a more robust feature space regardless of the classifier used.

An instance of the pipeline and an instance of each of the classifiers were trained and tested per participant. The FMG data from the training session was used to calibrate the pipeline. The data was used to train the FDA model, which transformed the data to fit the PCA model, which then transformed the data to train the UMAP method. Finally, training data transformed by the

pipeline was used to train the classifiers, whereas the test sessions' data was transformed by the pipeline and then used to evaluate each classifier's accuracy.

2.2.4 Statistical and Validation Methods

Three-way ANOVA was computed to show the impact of different factors on classification accuracy. The three independent variables were the classifier type, the test session order, and the data features, raw FMG vs. pipeline features. The classification accuracy was selected as the dependent variable. Post-hoc pairwise comparisons, Tukey's HSD (honestly significant difference) was conducted for independent variables for any significant effects. The significance level was set to $p\text{-value} = 0.05$.

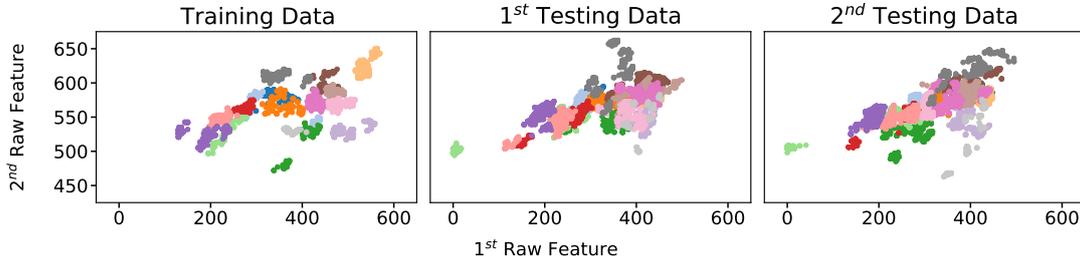
2.3 Results

The pipeline improved the separation between classes' data and reduced the variation within each class; thus, the classification performance significantly improved on the pipeline features. The transformation of the data throughout each step of the pipeline is visualized in Figure 2.3 while focusing and one gesture data in Figure 2.4. The mean accuracy results are reported in Figure 2.5, whereas per-class accuracy for several classifiers is depicted in confusion matrices in Figure 2.6. Furthermore, QDA performance is investigated for all participants in Figure 2.7.

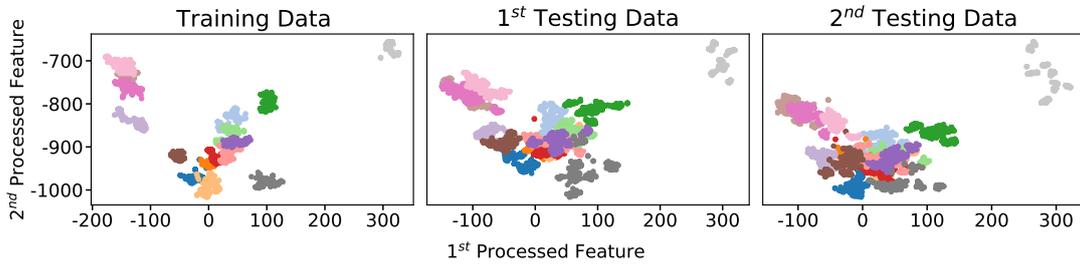
Figure 2.3 shows the comparison of the data distribution of participant 2 using the raw data (Figure 2.3a) and the data processed by the pipeline (Figure 2.3d), throughout the training and two testing sessions. In each plot, the x-axis and y-axis show the first two dimensions of the raw FMG data and the pipeline's features. The distributions of the gesture classes became easier to separate after the pipeline processing (Figure 2.3d) compared to raw FMG data in Figure 2.3a. On top of improving the separation between classes' data, the inter-class variance, the pipeline reduced the signal variance within a class, the intra-class variance.

In order to highlight the reduction of the intra-class variation, Figure 2.4 depicts only a single gesture data from participant 2 during the processing steps of the pipeline in the same manner as in Figure 2.3. Comparing Figure 2.4d to Figure 2.4a, we can see that the intra-

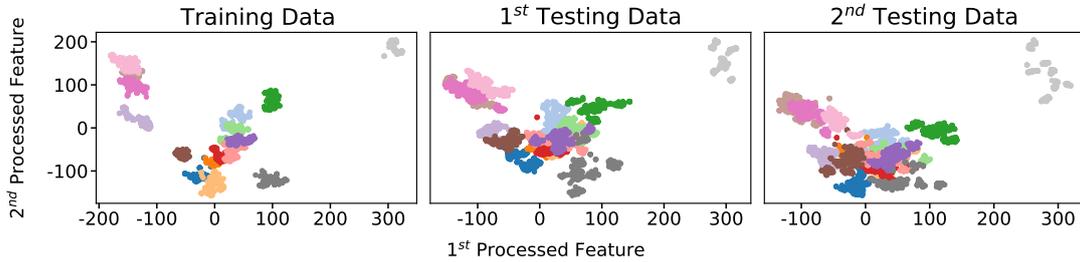
- Large Diameter Heavy Wrap (G1)
- Small Diameter Heavy Wrap (G2)
- Medium Wrap (G3)
- Adducted Thumb Wrap (G4)
- Light Tool (G5)
- Thumb + 4 Fingers Pinch (G6)
- Thumb + 3 Fingers Pinch (G7)
- Thumb + 2 Fingers Pinch (G8)
- Thumb + 1 Finger Pinch (G9)
- Disk Power Grasp (G10)
- Sphere Power Grasp (G11)
- Disk Precision Grasp (G12)
- Sphere Precision Grasp (G13)
- Tripod Grasp (G14)
- Push (G15)
- Lateral Pinch (G16)



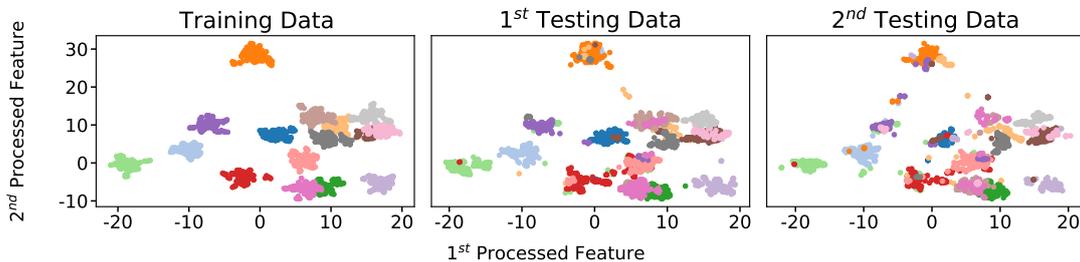
(a) Sessions data using raw features.



(b) Sessions data post applying FDA, the first step of the pipeline.



(c) Sessions data post applying PCA, the second step of the pipeline.



(d) Sessions data post applying UMAP, the final step of the pipeline.

Figure 2.3: Feature spaces after applying each pipeline step for all participant 2 data gestures, displaying the first two dimensions for each feature space.

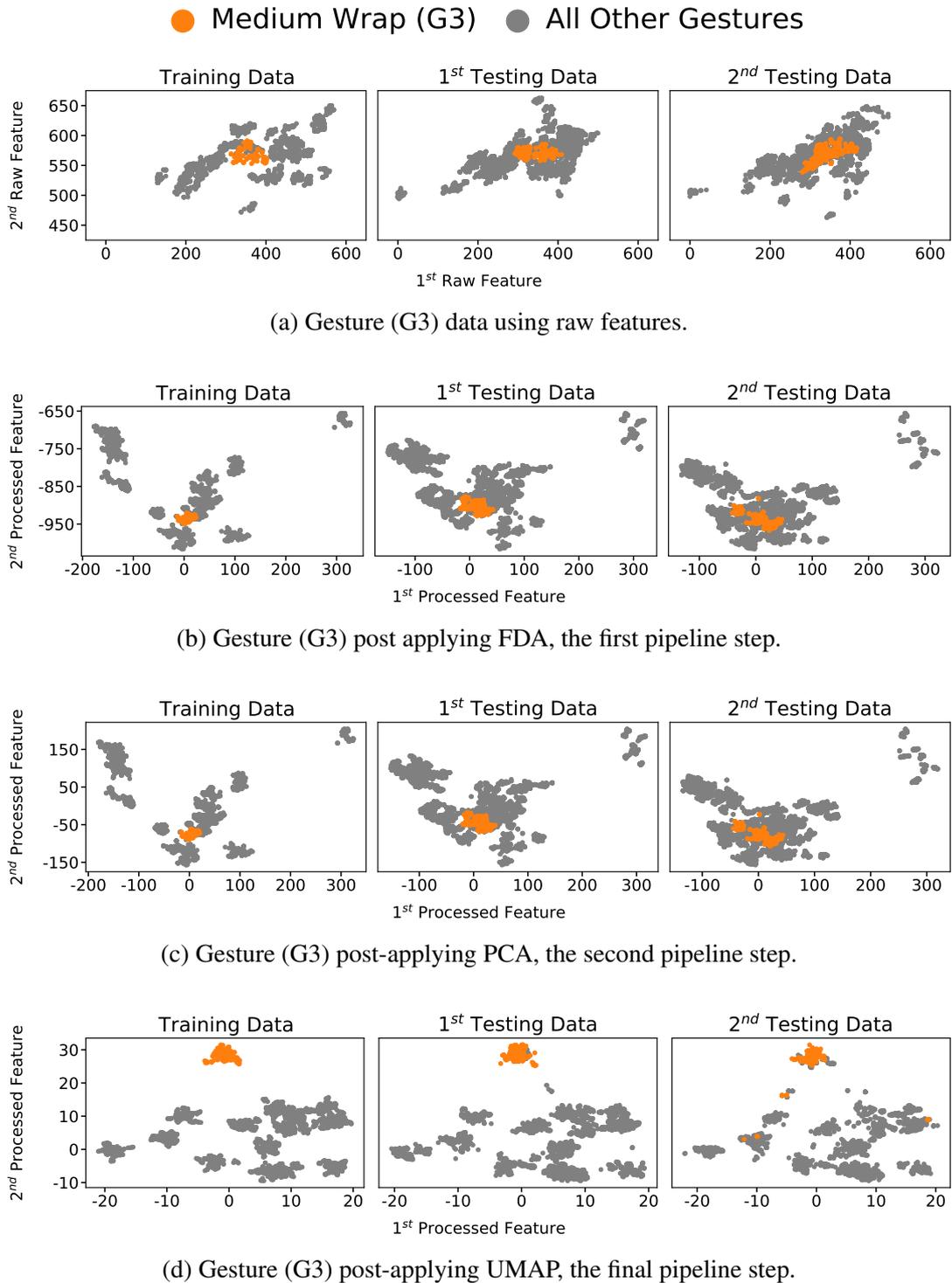


Figure 2.4: Feature spaces after applying each pipeline step, highlighting medium wrap gesture (G3) of participant 2 training data, showing the first two dimensions for each feature space.

class variance was reduced and the data points for the highlighted gesture data after pipeline processing clustered together.

Figure 2.5 shows the mean accuracies over all participants' data per test session using raw

features (Figure 2.5a) and pipeline features (Figure 2.5b). We can see that the accuracies have consistently improved for all classifiers after pipeline processing for both testing sessions.

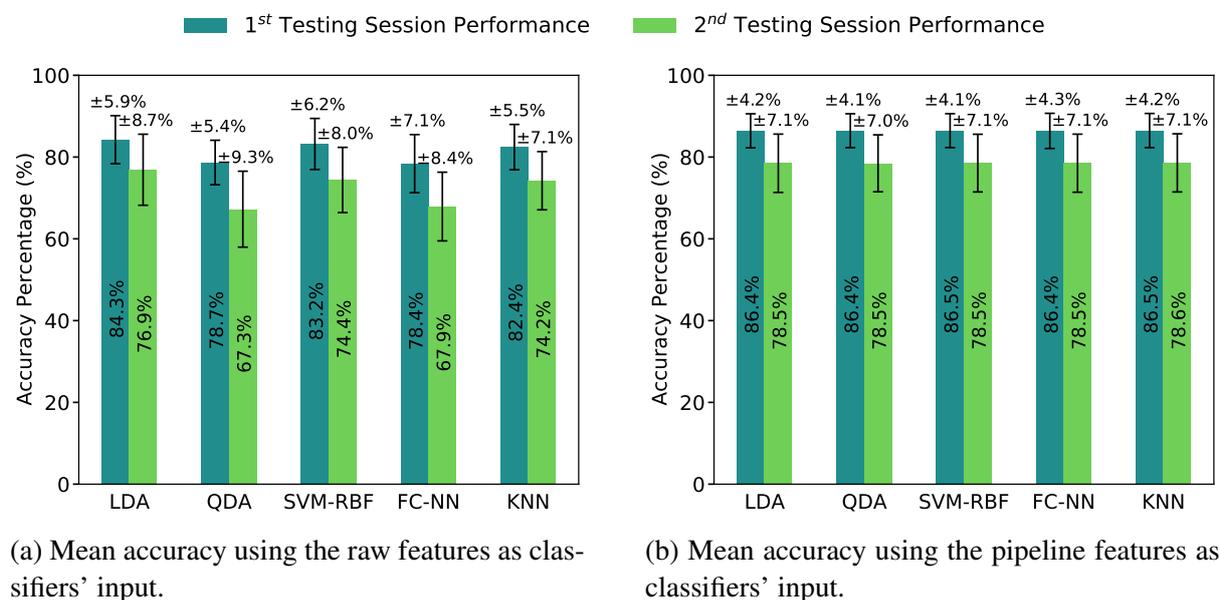


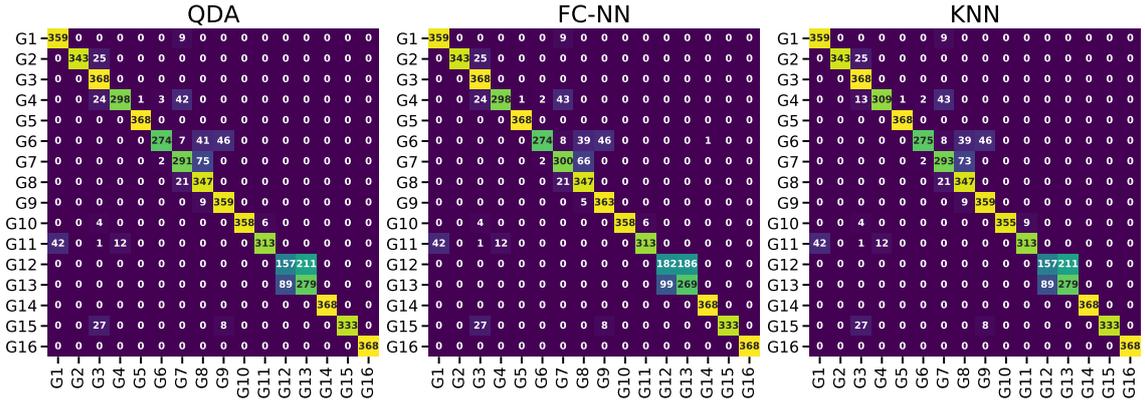
Figure 2.5: Classifiers' accuracy averaged over all participants for each test session using raw and pipeline features. Error bars resemble one standard deviation.

To investigate the cause of the accuracy similarity of the different classifiers, we present the per-class accuracies in confusion matrices in Figure 2.6. The confusion matrices are computed for 3 of the classifiers for participant two's first and second test sessions.

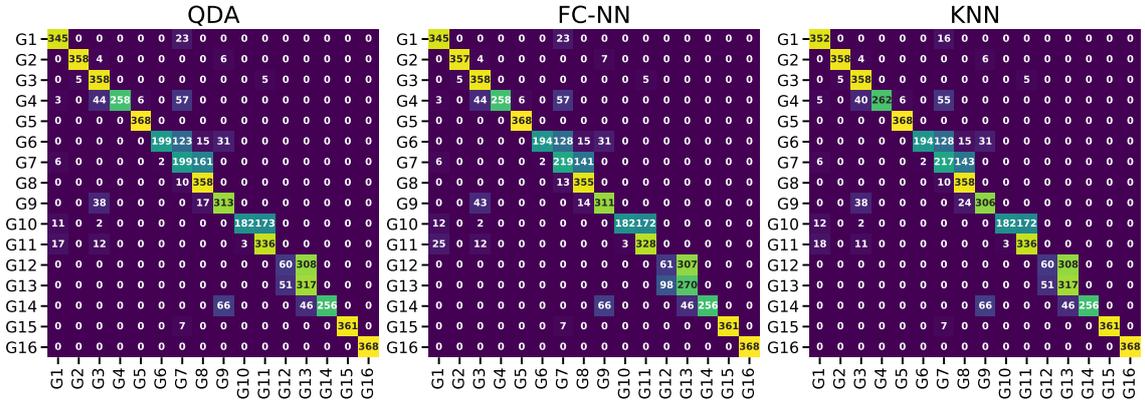
The confusion matrices for participant 2 in Figure 2.6 further support the similarity of results. For each test session in the figure, three different classifiers misclassify almost the same overlapping outliers that are hard to classify correctly, regardless of the model employed.

Among the five classifiers, the QDA classifier most benefited from applying the proposed pipeline to the data. Figure 2.7 shows each participant's QDA accuracy results for raw features and the proposed pipeline features, respectively. As shown in Figure 2.7, the pipeline improves QDA performance for all participants except for participant 4, whose second test session accuracy decreases. We conclude that this session is an outlier, given the consistent performance of the pipeline with all other test sessions.

The stochastic variance of FMG signals led to declined classification performance and massive variance in accuracy of different classifiers (Figure 2.5a) and different participants (Figure 2.7a). The pipeline reduced the variance above, making the accuracy difference in the test



(a) Confusion matrices using participant two's first test session.



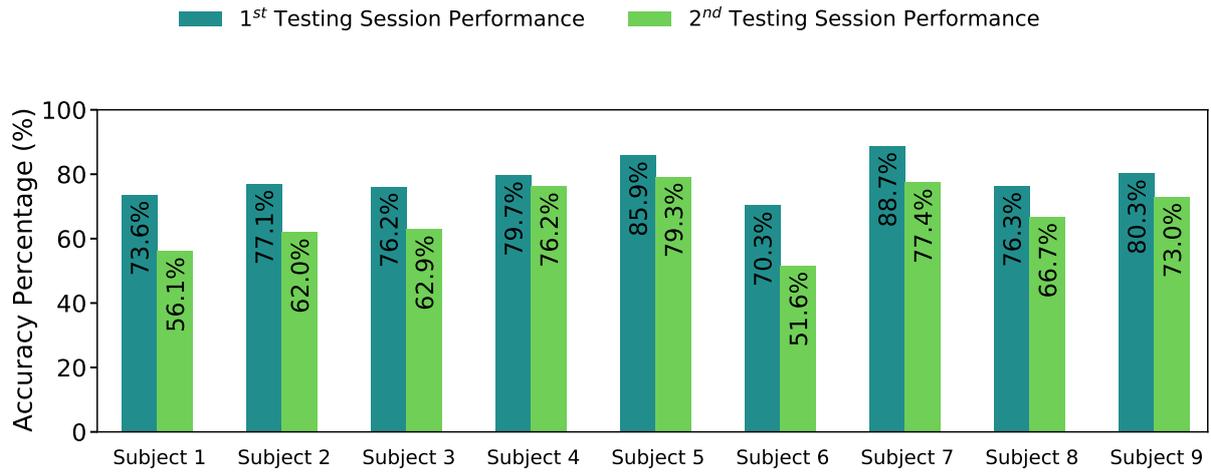
(b) Confusion matrices using participant two's second test session.

Figure 2.6: Confusion matrices for QDA, FC-NN, and KNN displaying per-class accuracy on the pipeline's features for participant two's test sessions.

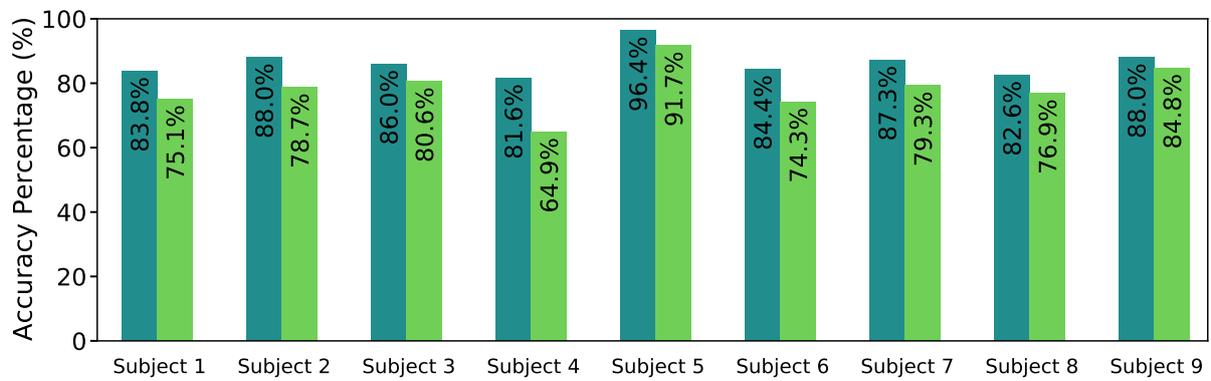
sessions near-constant in Figure 2.5b and reducing the variance for all classifiers for the nine participants in Figure 2.7b.

To investigate individual steps' impact on classification accuracy, we obtained the recognition results for applying each pipeline step separately in Table 2.3. These results explain the individual effects, depicting the transformation distribution at each step in Figure 2.3.

ANOVA results showed that the test session order ($F_{1,199} = 59.304, p \leq 0.0001$) and data feature space ($F_{1,199} = 29.690, p \leq 0.0001$) impact the recognition accuracy significantly; whereas the classifier type had marginal significance to the accuracy ($F_{4,199} = 2.373, p = 0.054$). There are no significant interactions between the independent variables, neither 2-way nor 3-way interactions. As the classifier type effect was insignificant, we further applied 1-way ANOVA to the results based on raw FMG and pipeline processed features, respectively, taking only classifier type as an independent factor.



(a) Quadratic Discriminant Analysis classifier accuracy results on raw features.



(b) Quadratic Discriminant Analysis accuracy results on the pipeline features.

Figure 2.7: Comparison of QDA accuracy using raw and pipeline features for each test session of all participants.

There was a significant effect on accuracy due to the classifier type using raw FMG features ($F_{4,99} = 3.083$, $p < 0.05$), but there was no significant difference between classifiers' after pipeline processing. Further post-hoc analysis on the raw FMG-based classification showed LDA had significantly higher accuracy than QDA ($p < 0.05$) and FC-NN ($p < 0.01$), respectively, affirming the results in Figure 2.5a. No other pairs of classifiers had a significant difference.

Table 2.3: Mean accuracy for the pipeline features and features of FDA, PCA, and UMAP individually as standalone preprocessing methods. The accuracy results are averaged over all participants and rounded to 1 decimal place.

Classifier	FDA Features		PCA Features		UMAP Features		Pipeline Features	
	Session 1	Session 2	Session 1	Session 2	Session 1	Session 2	Session 1	Session 2
LDA	86.5 %	77.4 %	84.3 %	76.9 %	81.5 %	72.5 %	86.4 %	78.5 %
QDA	79.4 %	67.7 %	78.8 %	67.3 %	81.5 %	72.5 %	86.4 %	78.5 %
SVM-RBF	83.2 %	75.4 %	85.3 %	76.0 %	81.5 %	72.5 %	86.5 %	78.5 %
FC-NN	82.1 %	72.8 %	79.6 %	68.4 %	81.5 %	72.6 %	86.4 %	78.5 %
KNN	86.5 %	78.4 %	82.4 %	74.2 %	81.6 %	72.5 %	86.5 %	78.6 %
Mean	83.6 %	74.2 %	82.3 %	72.5 %	81.5 %	72.5 %	86.4 %	78.5 %
Std Dev	±10.4 %	±12.9 %	±11.4 %	±13.8 %	±10.8 %	±12.1 %	±8.6 %	±11.0

2.4 Discussion

2.4.1 Gesture Recognition Enhancement

After applying the pipeline processing, the recognition accuracy improved significantly and consistently for all the classifiers for both test sessions, as shown in Figure 2.5b. From simple to non-linear classifiers, all five models obtained almost identical accuracy and standard deviations, regardless of their performances prior to the pipeline in (Figure 2.5a). The performance similarity illustrates that the models learned from the same underlying structure and that the pipeline produced reliable features, which enhanced all classifiers' accuracy by varying degrees.

The higher separability of the pipeline distribution classes, compared to raw FMG distribution, is supported by Figure 2.3d. Additionally, the misclassifications using the processed features were due to noise in the data, which can be seen in the remaining overlap between different classes. The processed data needed only the simplest of decision boundaries to obtain the highest recognition accuracy in the presence of noise; hence the classification accuracy became independent of the classifier's complexity.

Without pipeline preprocessing, there was significant variance in classification results for different classifiers, as shown in Table 2.3. Interestingly, simpler models, such as LDA, achieved higher accuracy using FDA and PCA features than more complex classifiers, such as FC-NN.

FDA or PCA features yielded results strongly dependent on the type of classifier, as seen

in Table 2.3. However, UMAP features appeared consistent and robust concerning classifier type but with lower mean accuracy. Therefore, the pipeline has all three methods' advantages, enabling all classifiers to achieve high and consistent accuracy. The comparison in Table 2.3 showcases the pipeline impact on the classification accuracy that one method could not obtain. The comparison results support the concept of tackling the FMG variance characteristics sequentially in a combined manner.

The pipeline effect on all participants' data and enhancement of all classifiers' accuracy reveals the benefit of incorporating it into forcemyography applications. The pipeline obtains a mean accuracy over all participants and classifiers of 86.4% with $\pm 8.6\%$ standard deviation for the first test session and 78.5% with $\pm 11.0\%$ standard deviation for the second test session. The pipeline results surpass all its individual processing steps in Table 2.3 while simultaneously reducing the variation in the accuracy results, which is most beneficial in hand gesture applications requiring prolonged FMG signal acquisition.

The classification accuracy using LDA based on raw FMG features in the present chapter was similar to the results from the previous study based on the same data set [18]. In contrast, the pipeline proposed in this chapter achieved more robust performance across different classification algorithms. Another similar state-of-the-art study by Anvaripour et al. [52] achieved almost 93% classification accuracy on six hand gestures using a similar FMG band on the forearm. They employed a feature extraction technique to enhance the classifier's performance.

Due to the difference between their study to ours, we need to compare the difficulty of the classification task. Their study's random chance classification is 16.66% for six gestures. In contrast, we have a random chance of correct classification of 6.25% for our 16 gestures. Thus our pipeline gives more accuracy gain to the classifiers. Furthermore, our testing condition was more rigorous. For instance, we trained the models using natural grasping force. We then evaluated them by two testing sessions' data using eight different grasping force levels. Due to the 16 grasps we use in our study, the performance achieved by our pipeline method is more beneficial to daily applications requiring several hand gestures.

It is worth pointing out that the pipeline could not entirely eliminate the accuracy decline caused by the stochastic variance of FMG signals. We observe this from Figure 2.5b by com-

paring the accuracy of test session 1 to test session 2; there were still significant differences between the two sessions even after pipeline processing. However, both testing sessions' results improved significantly, demonstrating that the pipeline works throughout multiple sessions.

2.4.2 Similarity of Processed Distributions

The stochastic FMG variance produced different distributions for different time sessions. In Figure 2.3, participant two's data is shown after each pipeline step. FDA managed to reduce the overlap to some extent, as seen in Figure 2.3b. Afterward, PCA removed the colinearity between dimensions. PCA can be seen as a combination of a translation and a rotation of the distribution. However, only the first two dimensions of the distribution are shown in Figure 2.3c.

Finally, UMAP in Figure 2.3d utilized these uncorrelated dimensions previously processed by both FDA and PCA to obtain the final output distribution of the pipeline and separate the classes significantly compared to the original distribution in Figure 2.3a.

The proposed pipeline increased the similarity of gesture data distributions between the training and test sessions. Moreover, the pipeline separated different classes' data and increased the inter-class variance, making the outliers in test sessions easier to be distinguished from the main gesture data distribution. In contrast, in Figure 2.3a, data distribution varied significantly from testing session 1 to testing session 2 using the raw features.

The reduction of intra-class variance is easier seen in Figure 2.4a with highlighted gesture (G3) data only, while Figure 2.4d illustrates gesture (G3) distribution on the pipeline's features. FDA and PCA processing in Figure 2.4b, 2.4c are not sufficient to obtain the same degree of separation, as the (G3) gesture data still suffers from considerable inter-class variance and overlap, whereas the same gesture is entirely separated using the entire pipeline as shown in Figure 2.4d. The change in data distributions in Figure 2.4a is known as covariate shift [52]. The covariate shift is from machine learning literature, describing the same phenomena as the inter-session variance mentioned in the FMG literature. The figure shows that the distribution for gesture (G3) has a similar mean and variance through different sessions. Notably, there were still class outliers; however, most points had the same distribution across sessions. In

addition, the gesture (G3) had a much higher intra-variance on raw features and overlapped with other gestures making its classification harder.

2.4.3 Limitation and Future Work

The pipeline method was tested on an FMG dataset from a study constrained to a lab setting and conducted in different time sessions with short breaks without taking off the FSR wristband during the break. Obtaining FMG data for longer intervals requires wearing the wristband for the entire collection protocol, as different wearings of the wristband introduce sensor misalignment as another source of variance. Future investigating of the processing pipeline effect for extended periods would be possible if the variance due to the shifts in sensors' positions is eliminated.

The feature space obtained by the processing pipeline cannot be interpreted as human-understood features. The interpretability could help understand the correlation between features obtained from muscle activations' and their corresponding hand gestures. For instance, the data of gestures (G5) and (G13) may seem unrelated according to FMG signals in Figure 2.3a. In contrast, the pipeline finds these two gestures very similar while clustering the data of each class in Figure 2.3d according to its processed features. Future investigation of this correlation could make the processed features human-understandable. A research field currently on the rise is explainable artificial intelligence, which can provide understandable insights into the features.

The proposed pipeline's output signal has several preferred qualities, such as class separation and variance reduction within a class. These properties allow the pipeline to be utilized in other applications like FMG signals analysis, interpretation, and discovery of latent patterns. Finally, this pipeline can enhance clustering techniques, given its class separability, as proved by the KNN improvement, which uses concepts similar to several clustering techniques.

2.5 Conclusion

The present work proposed a data processing pipeline to improve hand gesture recognition performance using machine learning to produce consistent data features. The results showed that the pipeline effectively maximized the inter-class signal variance and minimized the within-class variance, separating different classes into unique clusters. Thus, the pipeline improves the classification reliability and accuracy when using different classifiers. As shown in this study, the pipeline performance was not affected by the participant's data distribution and classifiers' types, as all participants' and classifiers' results were improved. This study reduces hand gesture recognition variation due to muscle activity and FMG stochastic variance, which has prominent potential to be used in more applications.

Results imply that the pipeline effect is not merely its features' consistency but also increased separation of classes, allowing simple classifiers to compete with complex ones. The similar performance of classification methods could enhance the inference experience for real-time gesture classification of FMG signals, as simpler models require less computation. Compared to the individual application of each of the pipeline components, the enhanced effect of adding them together as a whole pipeline demonstrated superiority in obtaining robust and higher accuracy results by deliberately tackling particular characteristics in the data.

2.6 Co-Authorship Statement

This chapter's results are obtained from the collaborative research of Mohammed Asfour, Dr. Xianta Jiang, and Dr. Carlo Menon. Mohammed Asfour is the primary investigator of the research. Dr. Jiang and Dr. Menon are senior researchers supervising and directing the research.

The authors roles in the research whose results are presented is as follows:

- **Conceptualization:** Mohammed Asfour, and Xianta Jiang
- **Data curation:** Xianta Jiang, and Carlo Menon
- **Formal analysis:** Mohammed Asfour
- **Funding acquisition:** Xianta Jiang

- **Investigation:** Mohammed Asfour
- **Methodology:** Mohammed Asfour
- **Supervision:** Xianta Jiang
- **Writing:** Mohammed Asfour
- **Review and editing:** Mohammed Asfour, and Xianta Jiang

Mohammed Asfour has obtained the consent of the two senior researchers, the co-authors, to publish the research and its findings as part of his thesis.

This chapter's research has been published as "Asfour, M.; Menon, C.; Jiang, X. A Machine Learning Processing Pipeline for Reliable Hand Gesture Classification of FMG Signals with Stochastic Variance. *Sensors* 2021, 21, 1504. <https://doi.org/10.3390/s21041504>" [1].

Chapter 3

Compatibility of Feature-Classifier Pairs for EMG Hand Gesture Recognition under Joint Processing Procedures

3.1 Introduction

Many fields, such as virtual reality (VR) [53], robotic arms control [54], or hand prosthesis [55] depend on hand gesture recognition. Its data could be visual using cameras [56], kinematic using motion sensors [57], or muscles signals detected by surface electromyography sensors (sEMG or EMG) [58] [59]. Out of them all, sEMG has significant advantages and is prominently used in medical devices, human-machine interaction [60] [61], and prosthesis control [62] [63], due to being a safe, easy to use, and noninvasive way to collect data.

The recent advancement in many fields made electromyography sensors more affordable and sensitive. For instance, Prakash et al. [64] developed an sEMG armband that acquired a 1.4 times Signal to Noise Ratio (SNR) on average and a 45% increase in sensitivity compared to a commercial EMG sensor. Whereas, a study showing the medical applications for sEMG by Dwivedi et al. [65] presented a virtual reality experience to help rehabilitate upper limb prostheses users. They evaluated their hypotheses about sEMG hand gesture classification with a Random Forest classifier after optimizing the feature window size, stride, and the number of

base models of the classifier. They successfully predicted the manipulated object movement in VR with up to 92% accuracy from the participants' EMG signals. In addition, their model predicted the gesture used with 83% classification accuracy.

Due to the importance of sEMG signals [66] in such applications, their classification and processing methods were extensively researched [67] [68], such as feature engineering [69]. Nonetheless, most literature focuses on a controlled environment that limits the generality of the results to other settings, such as the choice of the classifier. For instance, Phinyomark et al. [70] investigated different sEMG features, such as Mean Absolute Value (MAV) and Waveform Length (WL), and the resulting recognition accuracy. They concluded that some features were redundant and recommended others. However, they used only the linear discriminant analysis classifier (LDA) and a fixed feature window size, disregarding the classifier's complexity and bias. Thus, the features' results could not be extended to other classifiers.

As for using studies' findings in different applications, most studies choose a specific application, making their results inconsistent with others. For example, offline gesture recognition studies, with large window sizes, obtain around 95% recognition accuracy [71] [72], whereas real-time studies, with small window sizes, obtain around 80-85% accuracy [73]. Benalcazar et al. [74] proposed a multiple-stage model for the acquisition, preprocessing, feature engineering, classification, and postprocessing of EMG signals. Using the KNN classifier, they surpassed a commercial armband accuracy by 3% on five gestures. They used a 1-second window with a 250 ms stride because real-time classification has to be under 300 ms.

Hence, these distinct settings in the EMG studies present the necessity for an exhaustive study of the various EMG processing and classification methods with the fewest predetermined conditions while exploring processing factors, such as the feature window size and signal normalization range. Such an analysis would be a solid basis for EMG research points and applications regardless of their specific study settings, such as features, classifiers, or other processing methods.

A study with a similar aim was conducted by Mendes Junior et al. [75] in which they investigated multiple classifiers' performance across different sEMG features from the literature. They used feature selection to determine the optimal feature combination per classifier. More-

over, they investigated the impact of multiple dimensionality reduction techniques on classification. However, without a standard window size in literature, their 2000 ms window could require relatively large computational resources and cause delay for real-time gesture classification. This hurdle is shown by Smith et al. [25], who confirmed an inversely proportional relationship between recognition error and the window size. Thus, their results could not be generalized to other filtering window sizes.

This chapter proposes the existence of feature-classifier pairing compatibility that radically governs recognition performance. We test this compatibility with varying window sizes and normalization ranges to obtain optimal processing and classification settings for various research points and applications without bias. Secondly, the optimal window size for gesture recognition ranges beyond window sizes that can be used in real-time applications. As a final analysis point, we investigate if a signal normalization range that maintains signal polarity, including more information, is optimal for most feature-model pairs. From several normalization formulas [76] [77] [78], we use the task peak values as it suits the scope of gesture recognition.

By incorporating different scenarios, we remove any bias towards a selected application or circumstances; thus, the results become generalizable. We discuss the findings of the combinatorial settings and rank them by recognition accuracy. Consequently, we deduce from the proposed hypotheses that a compatible feature-model pair with a moderately large window will surpass any configuration for hand gesture recognition. Research that benefits from this analysis vary from real-time with a small window to offline recognition. Others vary from limited computation using linear models to higher-end systems with ensemble models without much delay.

3.2 Materials and Methods

3.2.1 Electromyography Data

The data is from a previous study [79] using a Noraxon Myosystem 1400 L acquisition device in Figure 3.1. The Interdisciplinary Committee on Ethics in Human Research (ICEHR) at the Memorial University of Newfoundland approved the secondary use of the data for our research

under application number 20230479-SC. The signals were obtained using eight bipolar sensors, 16 channels, at a sampling rate of 1000 Hz with an amplification gain of 500. The data was collected from 12 fully operational right-handed participants, six males, and six females. All participants signed a consent form approved by Simon Fraser University. Participants completed three label sets; each had 16 hand gestures, shown on a screen with the object and gesture to perform. Signal samples are shown in Figure 3.2. Further description of the gestures with images is in the original data collection study [79].

The data is broken down in Figure 3.3 as follows:

- **Sessions:** Two data collection sessions were conducted. In the first, electrodes were on the forearm of the participants, whereas in the second, they were on their wrists, providing different scenarios to test if our hypotheses can be generalized to multiple scenarios.
- **Label Groups:** For each session, signals of three sets of gestures were collected. The first is 16 hand grasps of different objects taken from Cutkosky's grasp taxonomy [80]. The second has 16 gestures from the American Sign Language (ASL) [81]. The third contains hand positions [82], such as pronation and supination.
- **Repetitions:** Each gesture in a label group was performed by every participant 5 times.

The dataset is chosen due to its variations to test hypotheses in multiple scenarios, such as different placement of sensors.



Figure 3.1: The MyoSystem 1400L EMG acquisition device.

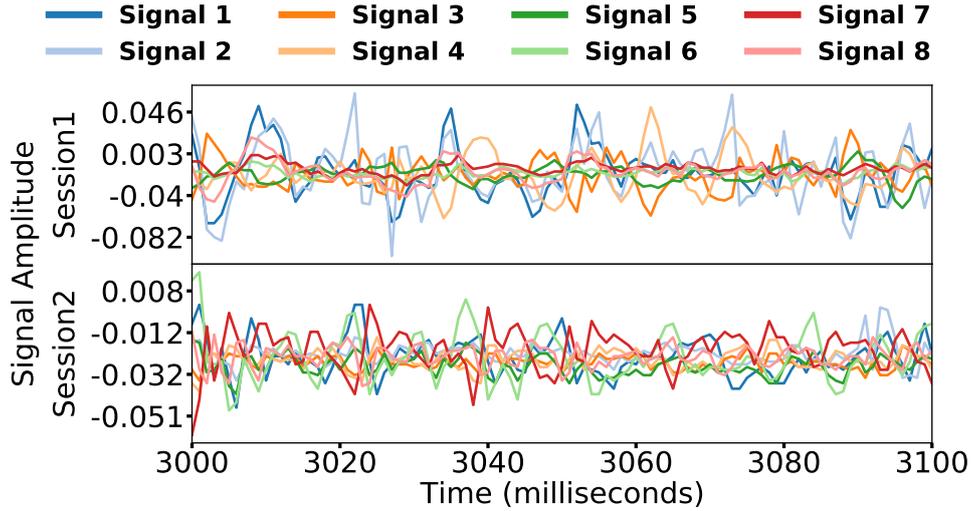


Figure 3.2: EMG samples acquired from participant 4.

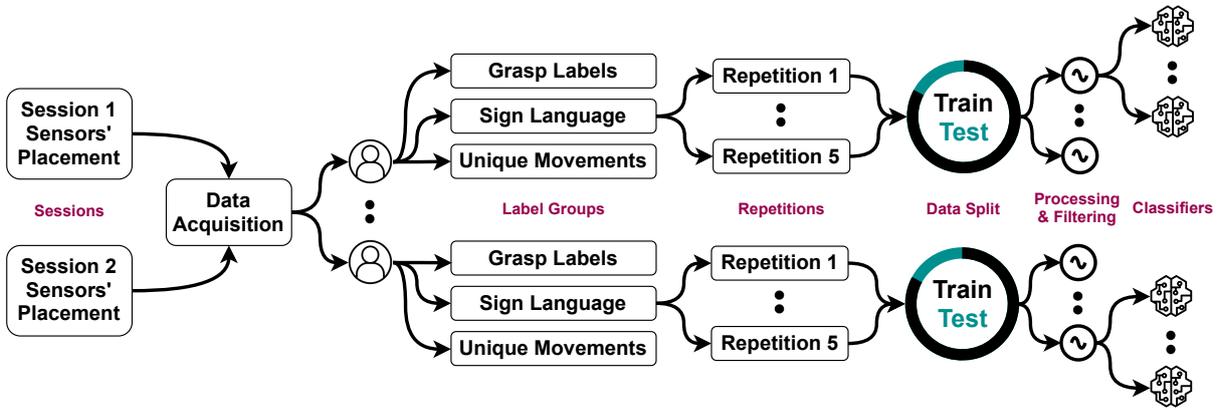


Figure 3.3: A diagram of the categories in the dataset.

3.2.2 Hypotheses Testing and Conducting Experiments

To study feature-model compatibility, we evaluate models of different theories and assumptions, as we believe their assumptions to be crucial to this relationship. We evaluate them using several sEMG features under a spectrum of feature window sizes and normalization ranges. The multiple test settings enable us to analyze the feature-model compatibility hypothesis without bias and find its optimal processing settings. This research is conducted via Python programming language.

Feature-Classifier Pairing Effect on Recognition Accuracy

We use various classifiers with different work theories to investigate the pairing impact thoroughly. The list of classifiers used and their theoretical assumptions are as follows:

- **Logistic Regression (LR)** [83] assumes the linearity between data features and classes' log probability.
- **Linear Discriminant Analysis (LDA)** [84] is another linear method. LDA assumes each class data is a Gaussian in feature space with the same covariance matrix.
- **Support Vector Machine with Linear Kernel (SVM-LIN)** [85] is a linear classifier that uses fringe points of each class to obtain the best linear boundary separating the classes.
- **Quadratic Discriminant Analysis (QDA)** [86] is a quadratic model with similar assumptions to LDA, except it assumes the uniqueness of each class's covariance matrix.
- **Naive Bayes (NB)** [87] model assumes the conditional independence between features given the data's class.
- **Decision Tree (DT)** [88] classifier sequentially splits the feature space based on learned thresholds.
- **K-Nearest Neighbors (KNN)** [89] stores the training data, without learning, to assign test data to the majority class of the nearest K neighbors in the feature space.
- **Random Forest (RF)** [90] is an ensemble of Decision Trees and is a critical classifier. In RF, a bagging technique, each tree is trained independently.
- **Gradient Boosting (GB)** [91] is a boosting ensemble of Decision Trees, using misclassified points of a DT to enhance the training of the next one.
- **Support Vector Machine with Radial Basis Kernel (SVM-RBF)** [85] has the same concept as SVM-LIN; however, it uses a kernel, a function resembling a metric in a hypothesized space instead of the dot product.

Deep Learning was excluded due to the dataset size. In each session, 400,000 samples were recorded per participant per label group [79]. Using the smallest window we investigate, 100 ms, yields 4,000 samples, which is extremely small for deep learning.

EMG Features

We pair the classifiers with commonly used twelve different sEMG features from literature [75]. Each feature is applied separately per sEMG electrode signal using a non-overlapping window as listed:

- **Root Mean Square (RMS)** [70] accumulates the square of the signals inside the sliding window from the equation

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

where N is the window size and x_i is the current reading.

- **Integrated EMG (IEMG)** [70] integrates the absolute values in the window to represent its total activation using

$$IEMG = \sum_{i=1}^N |x_i| \quad (3.2)$$

where N is the window size and x_i is the current reading.

- **Mean Absolute Value (MAV)** [70] computes the mean of the absolute values inside the sliding window using

$$MAV = \frac{1}{N} \sum_{i=1}^N |x_i| \quad (3.3)$$

where N is the window size and x_i is the current reading.

- **Waveform Length (WL)** [70] accumulates the signals in the window as its representation using the equation

$$WL = \sum_{i=1}^{N-1} |x_{i+1} - x_i| \quad (3.4)$$

where N is the window size and x_i and x_{i+1} are the current and next readings respectively.

- **Log Detector (LOG)** [70] applies to each sEMG sensor data for each window the following processing

$$LOG = \exp\left(\frac{1}{N} \sum_{i=1}^N \log(|x_i|)\right) \quad (3.5)$$

where N is the window size and x_i is the current reading.

- **Simple Square Integral (SSI)** [70] integrates the square values of the signals in each window using the formula

$$SSI = \sum_{i=1}^N x_i^2 \quad (3.6)$$

where N is the window size and x_i is the current reading.

- **Variance of EMG (VAR)** [70] computes the variance of each sEMG sensor data for each window using

$$VAR = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2 \quad (3.7)$$

where N is the window size, x_i is the current reading, and \bar{x} is the mean of the signals in the window.

- **Willison Amplitude (WA)** [70] computes how often two subsequent readings' difference exceeds a threshold using

$$WA = \sum_{i=1}^{N-1} f(|x_i - x_{i+1}|) \quad (3.8)$$

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

where N is the window size and x_i and x_{i+1} are the current and next readings. δ is the threshold value.

- **Slope Sign Change (SSC)** [70] computes how often the sEMG signal changes their sign using the equation

$$SSC = \sum_{i=2}^{N-1} f([x_i - x_{i-1}] \times [x_i - x_{i+1}]) \quad (3.9)$$

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

where N is window size and x_{i-1} , x_i , and x_{i+1} are previous, current and next readings. δ is threshold value.

- **Skewness (SKW)** computes the lack of symmetry in the sliding window data using the equation

$$SKW = \frac{\sum_{i=1}^N (x_i - \bar{x})^3 / N}{\delta^3} \quad (3.10)$$

where N is the window size, x_i is the current reading, \bar{x} and δ are the signals' mean and standard deviation.

- **Kurtosis (KURT)** computes the degree to which the data is heavy-tailed or light-tailed relative to a Gaussian using

$$KURT = \frac{\sum_{i=1}^N (x_i - \bar{x})^4 / N}{\delta^4} \quad (3.11)$$

where N is the window size, x_i is the current reading, \bar{x} and δ are the signals' mean and standard deviation.

- **Signal Histogram (HIST)** counts the signals in bins, frequency domain dimensions. For each bin, HIST counts a sensor's readings, and averages over all sensors using

$$HIST = \frac{1}{S} \sum_{j=1}^S \sum_{i=1}^N I(B_{lower} < x_{j,i} \leq B_{upper}) \quad (3.12)$$

where N is the window size, $x_{h,i}$ is the current reading of the j^{th} out of S sensors, B_{lower} and B_{upper} are the bin's bounds. HIST changes the number of data dimensions.

Influence of Window Size on feature-Classifier Pairing Recognition Accuracy

We investigate nine window sizes, 100, 250, 500, 750, 1000, 1250, 1500, 1750, and 2000 ms, with non-overlapping stride, to monitor if some feature-classifier pairs surpass others by altering the window size. The findings of this analysis are essential for the performance-delay balance of gesture recognition. We then rank the pairs by balanced accuracy.

Signal Normalization Range Optimality

We evaluate four normalization ranges as a preprocessing step. We computed the normalization parameters from training data only and used them on each participant’s training and test data individually.

Despite applying a single feature to the eight sEMG signals giving similar value ranges, we normalize each sensor’s signals per participant to compare results between participants’ different signal amplitudes. Because the training data values influence the classifiers’ performance, normalization ensures that the classifiers are not affected by varying signal amplitudes from training to testing for the same participant. This change is known as the inter-session variance for sEMG gesture recognition.

De Luca et al. [92] and Konrad et al. [93] explained that signal normalization is essential to make sEMG independent of unwanted characteristics and to compare their results between several users or variations in sensors placement. Besides, our study includes two different sensor placements among 12 participants. We experiment with three ranges that preserve signal polarity: $(-1, 1)$, $(-2, 2)$, and normalization-free raw signals. We add $(0, 1)$ normalization as a typical machine learning approach. We analyze these procedures using non-tunable feature-model pairs with 100, 500, and 1000 ms window sizes to remove any bias towards any single configuration.

3.2.3 Experiments’ Control Settings for Generalizable Results

For the conclusions to be valid and extendable, we have to guarantee that information is not transferred from the testing data to the models. We carefully formulate requirements and rules to test any hypothesis to hold this criterion.

Experimenting on a Subset of Participants and Label Groups

We only used three random participants’ data to conduct the experiments. Using only 3 participants’ data to dissect ensures that the study’s findings are clear of two main hindrances. The first predicament arises from using data from one participant, which might incorporate too many anomalies; thus, the conclusions can not be generalized to other participants. Alterna-

tively, selecting too many participants for analysis would jeopardize generalizing the findings to settings outside this study because it would be prone to bias toward these specific participants.

Further, we used only the labels from the first session of the grasp gestures group to examine the research hypotheses without learning from all labels' groups or sensors' placement. The whole dataset is used to validate the research's findings to check if the hypotheses are generalizable to other gesture types.

Data Splitting by Repetition

We split each participant's data by repetitions to ensure no data sharing between training and testing, as signals from the exact repetition would have similar characteristics. In all experiments, we use four training repetitions (80%) and one test repetition (20%) with cross-validation, changing the repetitions assignment per iteration.

For each EMG feature type, the participant's data consisted of 8 columns, their features, except for the HIST feature, in which the number of feature columns was equal to the number of bins. After data cleaning and removing transitions between gestures, rows were reduced from 400,000 to 240,000 for each participant per label set in each collection session. The number of rows after applying the features varied depending on the window size.

Training and Evaluation Classifiers per Participant

For each participant in this study, we use 12 features. Per each feature, we train and evaluate ten classification models using cross-validation. Each participant's classifiers were trained and evaluated on their training and test subsets without inter-participant testing, thus making them subject-specific.

We use the balanced accuracy metric as it is used for multi-class classification and handles class imbalance. The balanced accuracy metric weighs the accuracy of each class relative to its number of points. Each class's accuracy has an equal contribution to the total accuracy. We interchange the "balanced accuracy" and "accuracy" to mean the balanced accuracy metric throughout this chapter.

Table 3.1: Values used in grid search for classifiers hyperparameters.

Classifier	Hyperparameter	Grid Values
SVM-LIN	C	0.1, 1, 5, 25, 45, 65, 85, 105, 125, 145
DT	Pruning Coeff Split Min Samples	0.0, 0.01, 0.02, 0.03, 0.04, 0.05 5, 10, 15
KNN	Distance Metric Neighbors Weights Neighbors (K)	Minkowski, Euclidean Uniform, Distance 5, 10, 15
RF	Pruning Coeff. # Base Models	0.0, 0.01, 0.02, 0.03, 0.04, 0.05 25, 50
GB	Pruning Coeff. # Base Models	0.0, 0.0025, 0.005, 0.0075, 0.01, 0.0125 25, 50
SVM-RBF	C	0.1, 1, 10, 20, 30, 40, 50, 60, 70, 80, 90

Classifiers' Hyperparameters Tuning

We perform grid search cross-validation for six classifiers with hyperparameters to optimize them. The grid is in Table 3.1.

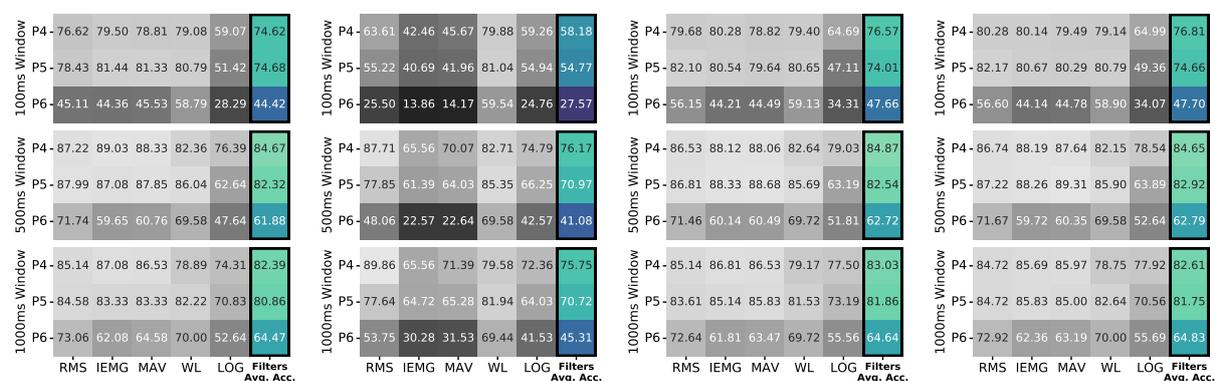
We started with evaluating a reasonable value for each hyperparameter and then evaluated the effect of its increment or decrement. We iterated the increments and decrements until reaching upper and lower limits, beyond which the classifier's performance degraded. We then generated a range of values between the upper and lower limits.

We apply the grid search with cross-validation with each feature to yield optimized hyperparameters specific to each feature-model pair. The search ensures that the pair results are objective, as each model was tuned specifically for the feature used. Due to the number of hyperparameters investigated, we cannot cover them all. For instance, the kernel scale for SVM-RBF, for which we use the fixed value of $\frac{1}{n_{features} * VAR}$.

3.3 Results

3.3.1 EMG Signal Normalization

We test the normalization firstly as a preprocessing step of EMG signals. In Figure 3.4, we apply four normalization ranges, each per subfigure, to raw sEMG data, followed by applying the RMS, IEMG, MAV, WL, and LOG features, on the x-axis, with varying window sizes, as separate rows, to check if the effect is feature-independent without bias to a specific configuration. Figure 3.4 shows the averaged accuracy of non-tunable classifiers, LR, LDA, and QDA, to judge the normalization objectively. As shown in Figure 3.4, (-1, 1) normalization gives the best results; thus, it is applied to process the data before the coming investigation points.



(a) No normalization. (b) (0, 1) normalization. (c) (-1, 1) normalization. (d) (-2, 2) normalization.

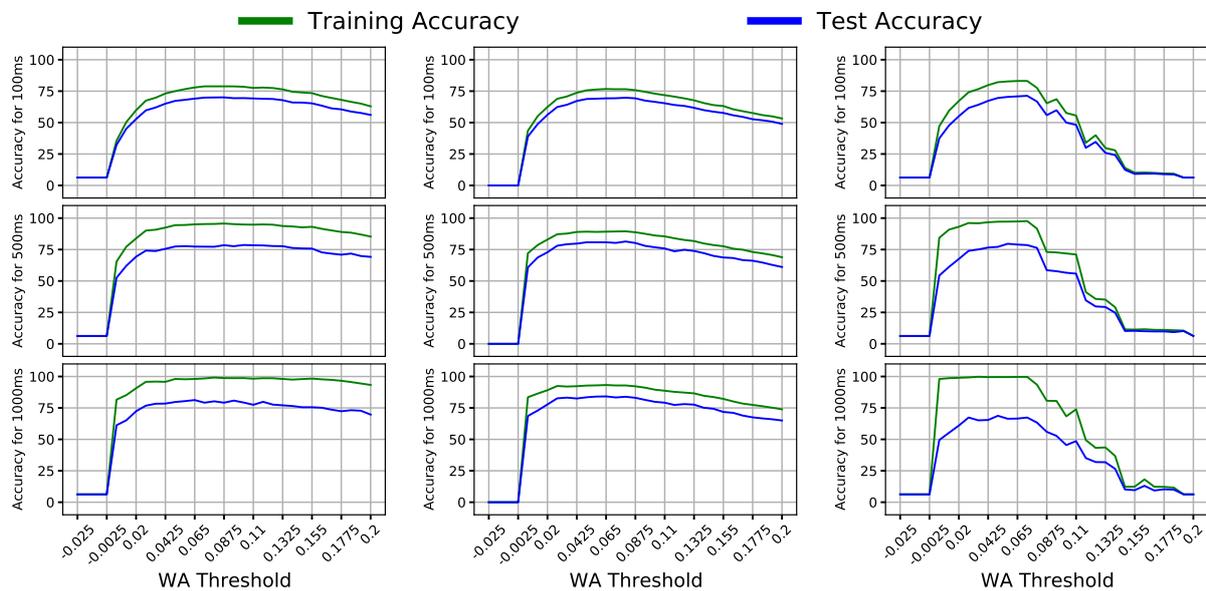
Figure 3.4: Accuracy results of Logistic Regression (LR), Linear Discriminant Analysis (LDA), and Quadratic Discriminant Analysis (QDA) averaged on participants 4, 5, and 6 data using no-tuning features and varying window size.

3.3.2 Grid Search for tunable Features and Classifiers

We tune the WA, SSC, and HIST features to optimize them first. Figure 3.5 and Figure 3.6 show a classifier per column and window sizes as rows. The x-axis contains the threshold values, whereas the y-axis resembles the accuracy.

Notably, Figure 3.5 shows a specific threshold range with the best accuracy across all classifier-window configurations. Therefore, the peak of this range is the optimal threshold for WA for this dataset. Similarly, SSC experiences a shared range of threshold values whose peak is chosen as the optimal threshold in Figure 3.6. The optimal values for WA, SSC, and

HIST are 0.065, 0.0066, and 30, respectively, and are used in the following experiments. We report the accuracy of two linear models, LR and LDA, and a quadratic model, QDA, to include the models' different assumptions and varying complexity.



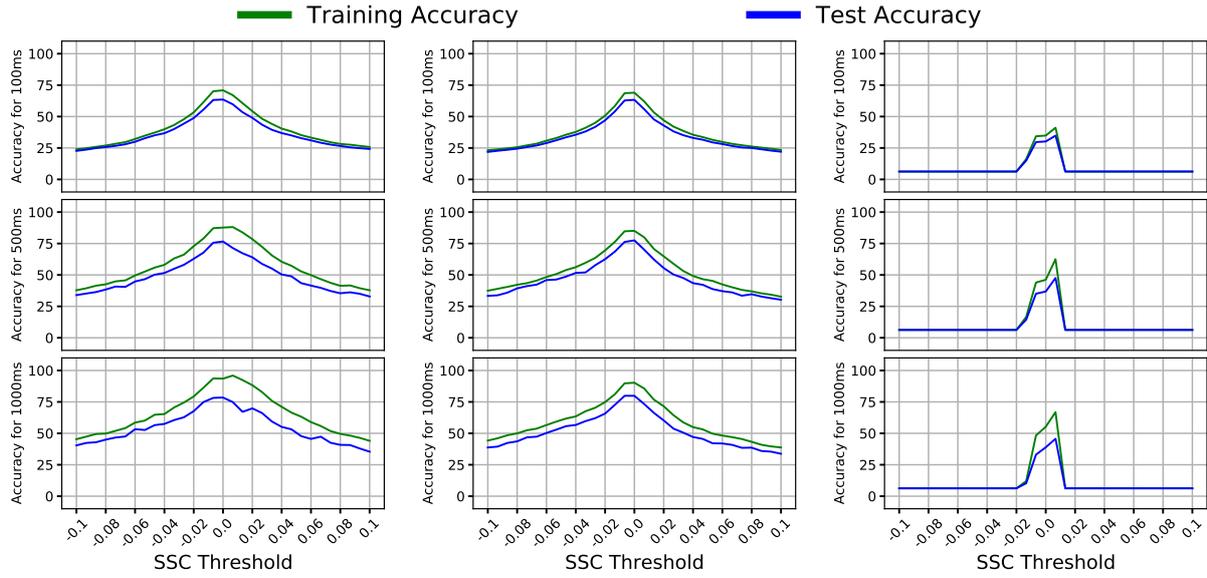
(a) Using Logistic Regression. (b) Using Linear Discriminant Analysis. (c) Using Quadratic Discriminant Analysis.

Figure 3.5: Accuracy results of Logistic Regression (LR), Linear Discriminant Analysis (LDA), and Quadratic Discriminant Analysis (QDA) with multiple WA thresholds on varying window sizes averaged on EMG data of participants 4, 5, and 6.

The grid search yields the optimal hyperparameters for models in Table 3.2 for each pair to tune each model to its paired feature. The grid results show that the optimal hyperparameters are the same regardless of the feature and window size. Hence, these values tuned the models to the recognition task, not towards a specific setting, feature, or window size.

3.3.3 Processing Window Size Impact on Recognition

From Figure 3.5 and Figure 3.6, we notice that window expansion notably enhances accuracy, aligning with our hypothesis. To investigate if expanding the window benefits the feature-classifier pairings, we illustrate the mean classification accuracy for the three participants. Figure 3.7 shows the mean recognition accuracy per each model-feature pair with window sizes of 100, 250, 500, 750, 1000, 1250, 1500, 1750, and 2000. The optimal window size in Figure 3.7 is 1250ms; thus, it is used in future experiments.



(a) Using Logistic Regression. (b) Using Linear Discriminant Analysis. (c) Using Quadratic Discriminant Analysis.

Figure 3.6: Accuracy results of Logistic Regression (LR), Linear Discriminant Analysis (LDA), and Quadratic Discriminant Analysis (QDA) with multiple SSC thresholds on varying window sizes averaged on EMG data of participants 4, 5, and 6.

Table 3.2: Classifiers' optimal hyperparameters after performing grid search using sEMG data of participants 4, 5, and 6.

Classifier	Hyperparameters
SVM-LIN	$C = 80$ (for Win=100), 85 (for other Win Sizes)
DT	Pruning Coeff = 0; Split Min Samples = 5
KNN	Distance Metric = Minkowski Neighbors Weights = Uniform; Neighbors (K) = 5
RF	Pruning Coeff = 0; # Base Models = 50
GB	Pruning Coeff = 0; # Base Models = 50
SVM-RBF	$C = 90$

3.3.4 Feature-Classifier Compatibility and Accuracy Ranking

We finally evaluate the accuracy of all feature-classifier pairs, averaged over all participants' data from the first session of the grasp labels using the optimal window size of 1250 ms. Distributions of accuracy results are shown in Figure 3.8

Figure 3.8 is summarized in Figure 3.9, whose first column contains the descending order

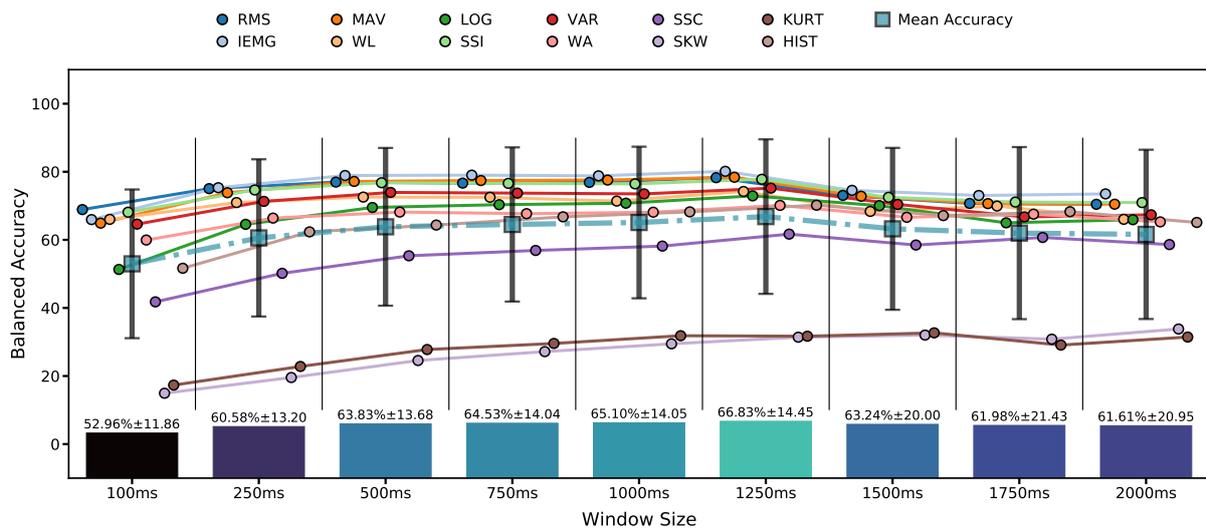


Figure 3.7: Features test recognition accuracy for different window sizes of all classifiers averaged on sEMG data of participants 4, 5, and 6.

of the best features on the optimal window of 1250 ms from top to bottom. The models are ranked descendingly from left to right for each feature, containing each feature-model pair's accuracy. Figure 3.9 shows IEMG and RMS features at 79.0% and 78.1% accuracy without threshold tuning, performing the best on average. At the same time, SKW and KURT are much less efficient regardless of the classifier, averaging around 32.1% and 31.9%, respectively.

The results in Figure 3.8 are also summarized in Figure 3.10 with the descending order of the models on the 1250 ms optimal window in the left-most column from top to bottom. The features are ranked, descendingly, from left to right for each model, showing the same pairs but using the model as the primary index. The Random Forest model is the best for the grasp group with 74.1% accuracy averaged on all features, yet HIST-LDA is the top with 88.63%.

Finally, we investigate if the results generalize for the sign language group, the hand movements group, and the other sensors' placement. Figure 3.11 contains the mean accuracy of the features for participants' signals from all sessions, and sensors' placement, of all label groups. Most features enable the classifier to perform well, except with SKW and KURT features. We notice that tunable features give worse accuracy than others, except for the HIST feature, yet, it has the most outlier results.

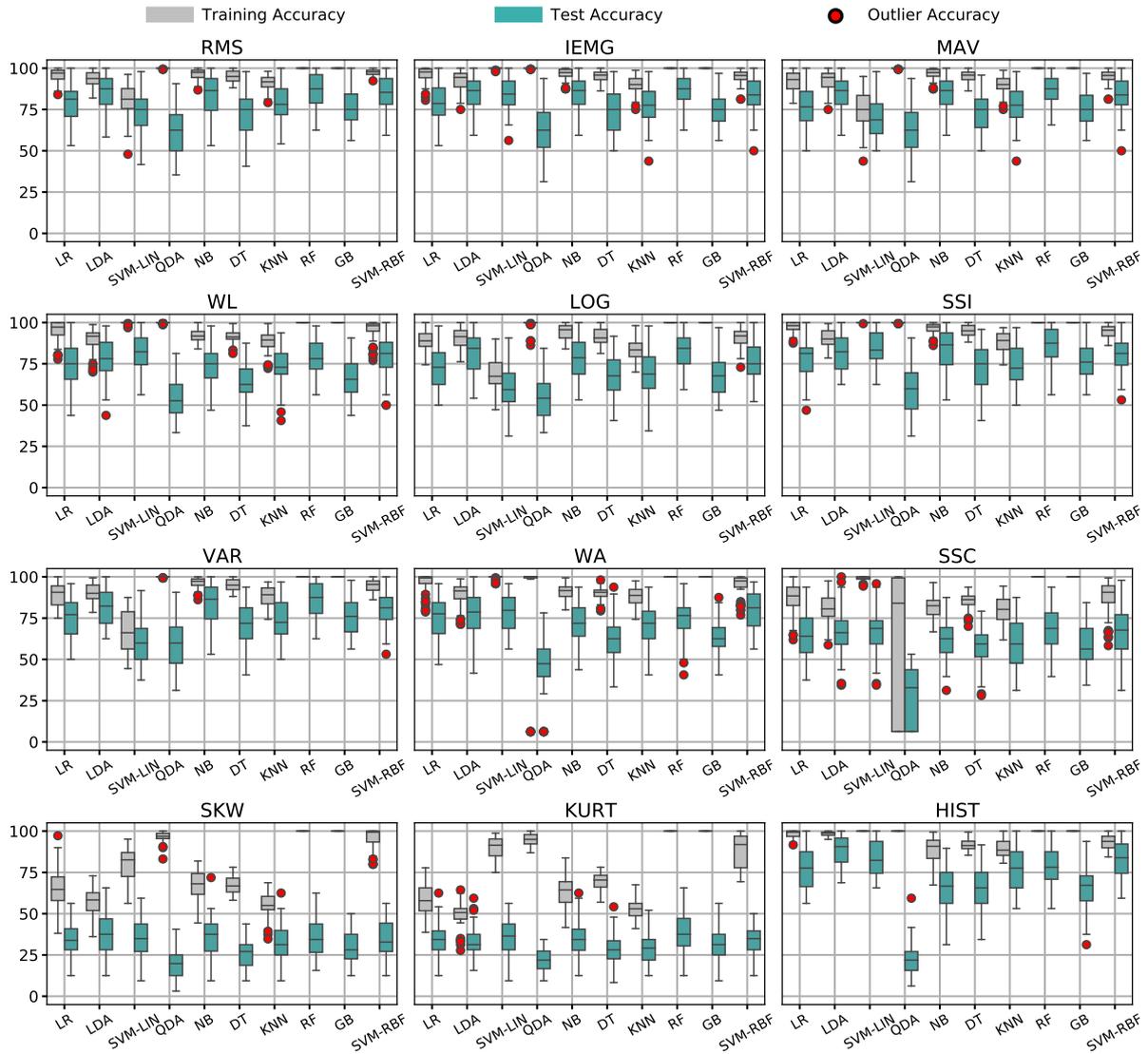


Figure 3.8: Distribution of accuracy results for all feature-model pairs for grasp label group averaged on all participants' data from the first session.

3.4 Discussion

The feature-classifier ranking in Figure 3.9 supports our primary hypothesis that the prominence of the pair compatibility on the recognition accuracy regardless of the models' complexity, as model ranking changes per feature. The SVM-LIN best witnesses this change, as it is the best model for the WL feature despite performing poorly on several features. We also find that the optimality of the 1250 ms window in Figure 3.7 proves the need for relatively large windows to include sufficient information for optimal performance as proposed. As for the normalization, Fig. 3.4 implies that the (-1, 1) range is sufficient for the best performance with no substantial enhancement by expanding the range.

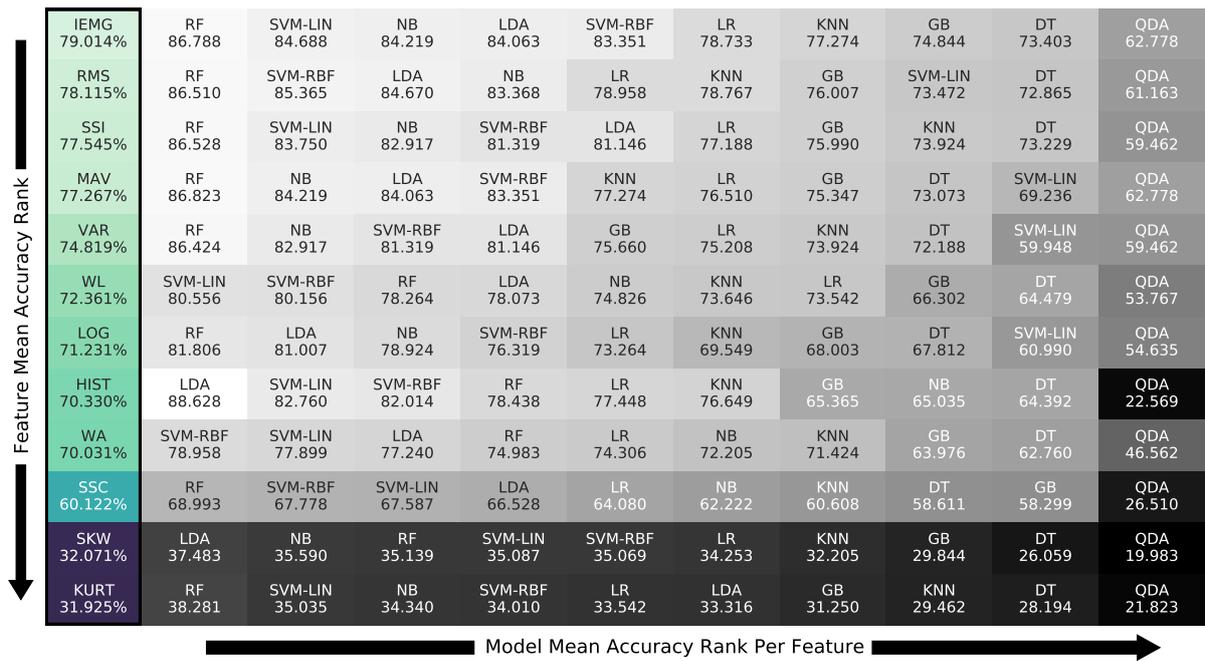


Figure 3.9: Feature accuracy ranking on the 1250 ms window with classifiers' sub-ranking averaged on all participants' grasp data.

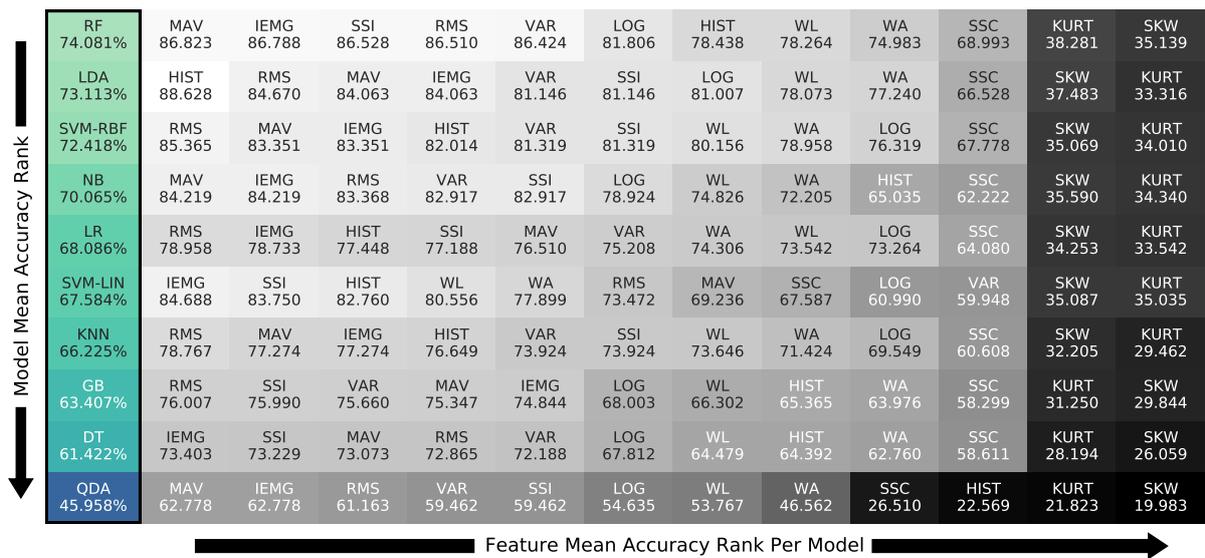
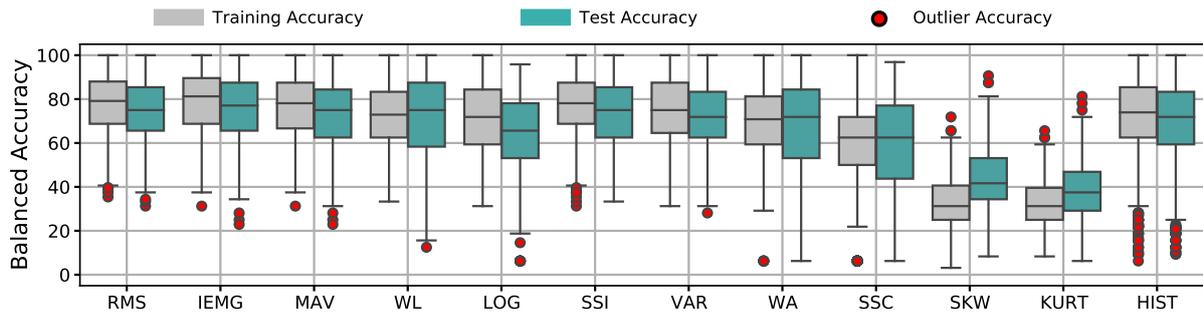
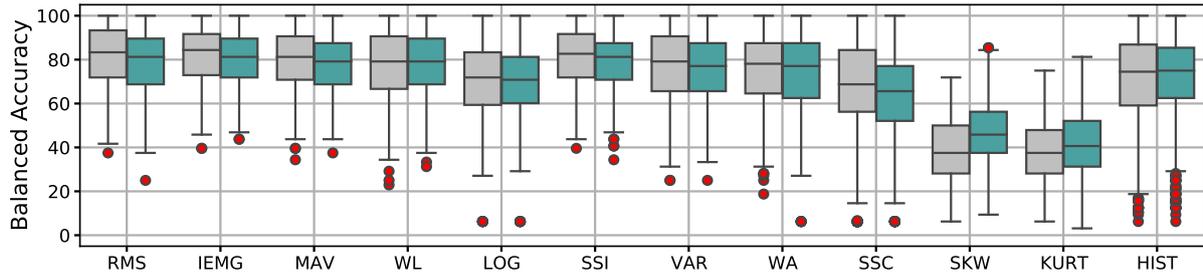


Figure 3.10: Classifier accuracy ranking on the 1250 ms window with features' sub-ranking averaged on all participants' grasp data.

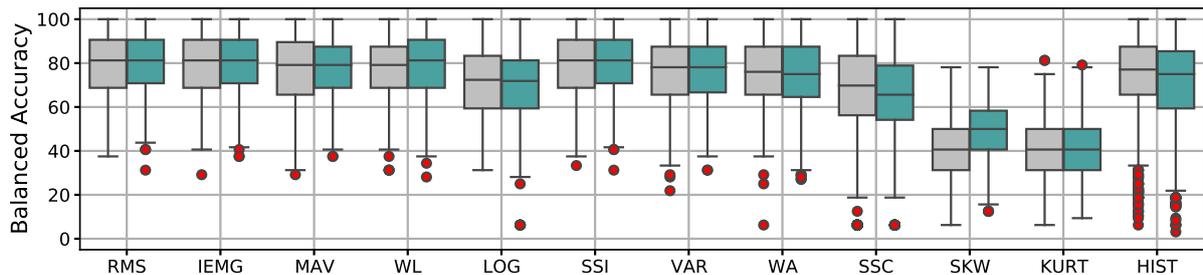
As the results show the impact of the proposed feature-model compatibility, we provide the ranking of the pairs in Figure 3.9 and Figure 3.10 to guide future research. The compatibility ranking reveals that simpler classifiers, such as LDA, are on par with ensemble models, such as RF, on particular features, such as HIST, RMS, and MAV. Therefore, to deliver accurate recognition with low computation, thus it is crucial to choose the appropriate pair for sEMG



(a) Distribution of accuracy on grasp label group.



(b) Distribution of accuracy on sign language group.



(c) Distribution of accuracy on unique movements label group.

Figure 3.11: Accuracy distribution per feature on Grasp Gestures, Sign Language Gestures, and Unique Movements Gestures averaged on all classifiers and all participants.

gesture classification. For instance, our ranking finds the Willison Amplitude (WL) feature as the most suitable for linear classifiers, as Support Vector Machine with Linear Kernel (SVM-LIN) outperforms all, including ensembles. In contrast, SVM-LIN performs erroneously on the VAR feature, whereas other linear models surpass the Gradient Boosting (GB) ensemble. We notice general behaviors of particular features or classifiers from their pairings' performances.

Focusing on models in Figure 3.9, the Random Forest (RF) is the top classifier with a small performance variability compared to others, as noted in Figure 3.8. In contrast, Quadratic Discriminant Analysis (QDA) performs worst with all paired features, implying that the theoretical assumptions of QDA are incompatible with the classification task at hand. This claim is backed

by the HIST feature's high to satisfactory results in Figure 3.9 on all classifiers except for QDA. These comparisons prove that the classifier's theoretical assumptions, known as theoretical bias, and its compatibility with the feature play a more significant role in the recognition process than its complexity. This is evident by the ensemble models surpassing linear ones for some features while failing for others.

Concentrating on the features in the ranking, we find that IEMG, RMS, MAV, and SSI are the top-performing ones unconditionally, regardless of the window. Thus, these features are the best choice regardless of the application type, real-time or offline recognition. Further, we note that SKW, KURT, and SSC, mainly tunable features, offer poor performance independent of the window. Their subpar performance is recorded in Figure 3.7 despite giving an acceptable performance for the three subjects' data used for tuning. Such results indicate that the tuned values do not generalize to the other participants. Figure 3.8 results concur with this observation, as all models perform accurately on all pairings but for SSC, SKW, and KURT features. These three features have low recognition accuracy and a distinctive overfitting issue. We infer from these two notes that these three features are signal-dependent in gesture recognition and must be tuned per user regardless of the classifier or feature window. Despite the top-performing HIST feature, it yields significant out-of-distribution results when generalized to other participants. HIST accuracy results are in Figure 3.11, confirming the need for user-specific feature tuning. We only report 12 commonly used features from the time domain, yet, more features, such as auto-regressive features, can be investigated in future work. Moreover, our research investigates single feature-classifier compatibility. However, feature selection and combination's effect on compatibility could be a future research topic, using this study as a basis.

Regarding our assumption about the feature window, Figure 3.7 indicates that the 1250 ms window is optimal for all investigated features. The result supports our assumption that the optimal window for classifying this number of labels, 16 gestures, is above the range of real-time applications to incorporate more information from signals for recognition. Our results agree with other findings in the literature [25]; however, the optimal window depends on the classification's difficulty, which corresponds to the number of labels of 16 hand gestures. Thus we analytically prove that a small window size does not contain enough data to achieve optimal

performance. Furthermore, overextending the window in Figure 3.7 can have counter outcomes, permitting signal noise to deteriorate the recognition performance throughout diverse control settings. However, a reasonable window of 250 ms reaches acceptable accuracy for real-time applications, forfeiting around 7% accuracy from the optimal window performance. This trade-off is a known accuracy-delay paradigm in sEMG gesture recognition literature and has been investigated under specific environment controls [25]. However, our study confirms it with various classifiers, features, and window sizes, thus having nearly no bias toward any specific model or feature. Increasing the window size removes most disparities between the features, making the feature choice less meaningful. The windows' overlap and the window stride are not covered in this study, which could be analyzed in the future.

As we hypothesized, $(-1, 1)$ normalization gives superior results to non-normalized features in Figure 3.4, in which $(0, 1)$ normalization gives a worse performance. From this remark, we deduce that the signals' polarity plays a critical role in gesture recognition. We confirm this by the higher performance of all procedures with negative to positive ranges than 0 to 1 normalization. On top of that, equalizing the upper and lower limits, -1 to $+1$, performed better than non-normalized signals. However, expanding the range to $(-2, 2)$ did not notably impact the accuracy. Thus, a basic polarity-preserving signal normalization that balances the range limits is optimal regardless of the feature-classifier pair.

Experiments yield equivalent results for different sensor placement and gestures, such as sign language and unique movements in Figure 3.11. Thus, our ranking and recommendations are generalizable to other recognition tasks and armband wearings.

Gesture classification applications treat false positives and false negatives errors equally; hence, the balanced accuracy metric is used throughout the literature and our study. Despite that, statistical analysis and other metrics might give an insight into the feature-classifier relationship as a future point which is excluded due to the number of points investigated. Furthermore, investigating the compatibility behavior differences among males and females or by age could give future insights into it.

3.5 Conclusion

This chapter explores the effect of pairing features and classification models on gesture recognition while influenced by combined processing methods, such as normalization ranges and window sizes. We provide a feature-classifier pairing ranking, the primary determinant of the recognition performance, as a guideline for future sEMG research to choose the appropriate pair for the research point or application. The ranking gives the best pairing for an application's pre-chosen feature or classifier. It also gives comparable feature-classifier pairs to the chosen ones with simpler classifiers. These advantages benefit any gesture recognition application, especially those with limited computational capabilities.

The compatibility's prominence is evident by the SVM-LIN achieving the best on the WL feature, 80.56% accuracy, and the second-worst on the MAV feature, 69.24% accuracy. The ranking shows linear models compete with complex ones on specific features, as HIST-LDA feature-model pair performs best with 88.63% accuracy for 16 gestures. Thus, the pair can perform high-accuracy gesture recognition with low computation, which implies that the models' complexity in gesture recognition is not as effective as believed. We present the optimal hyperparameters under different scenarios for future research on various conditions.

Secondly, the 1250 ms window is optimal for 16 hand gestures in this recognition task. Raising or lowering its size reduces classification accuracy. Nonetheless, real-time applications that require window sizes less than 300 ms benefit from the fact that the 250 ms window provides acceptable accuracy with a 7% reduction from optimal performance. We further conclude that the inequalities between feature performances fade for larger window sizes.

As for signal normalization, the standard normalization, -1 to +1, is unconditionally optimal for any gesture recognition task. Other ranges are less or equally valuable, particularly ones removing the signal polarity, which yield notably worse recognition accuracy.

3.6 Co-Authorship Statement

This chapter's results are obtained from the collaborative research of Mohammed Asfour, Dr. Xianta Jiang, and Dr. Carlo Menon. Mohammed Asfour is the primary investigator of the re-

search. Dr. Jiang and Dr. Menon are senior researchers supervising and directing the research.

The authors roles in the research whose results are presented is as follows:

- **Conceptualization:** Mohammed Asfour, and Xianta Jiang
- **Data curation:** Xianta Jiang, and Carlo Menon
- **Formal analysis:** Mohammed Asfour
- **Funding acquisition:** Xianta Jiang
- **Investigation:** Mohammed Asfour
- **Methodology:** Mohammed Asfour
- **Supervision:** Xianta Jiang
- **Writing:** Mohammed Asfour
- **Review and editing:** Mohammed Asfour, and Xianta Jiang

Mohammed Asfour has obtained the consent of the two senior researchers, the co-authors, to publish the research and its findings as part of his thesis.

This chapter's research has been published as "Asfour, M.; Menon, C.; Jiang, X. Feature–Classifier Pairing Compatibility for sEMG Signals in Hand Gesture Recognition under Joint Effects of Processing Procedures. *Bioengineering* 2022, 9, 634.

<https://doi.org/10.3390/bioengineering9110634>" [2].

Chapter 4

Conclusion

This work presents methods to enhance hand gesture classification of electromyography (EMG) and (FMG). By introducing a machine learning pipeline that yields robust features in Chapter 2 and recommending a feature-classifier pairing ranking, in Chapter 3, EMG and FMG gesture recognition common errors are minimized.

The machine learning pipeline processed signals into robust features, increasing the classification consistency across different classifiers for FMG signals. It partially limits intra-subject recognition error, reducing the accuracy gap between testing sessions. The classifiers' performance was inconsistent as different gestures' data on the raw FMG features was heavily overlapping. Linear discriminant analysis (LDA) performed better than most other classification methods, including highly non-linear ones; however, we show that it is due to randomness rather than learning a reasonable decision boundary. We find that quadratic discriminant analysis (QDA) performs worse than LDA, despite being similar but with higher complexity. The results illustrate an intra-user inconsistency that comprises the recognition performance, depending on randomness rather than the data.

The proposed pipeline in Chapter 2 composes of Fisher's discriminant analysis (FDA), principal component analysis (PCA), and uniform manifold approximation and projection (UMAP), in that order. FDA is used to transform the raw feature space linearly into the best separation between signals of different gestures, making their correct classification more feasible. PCA removes any correlation between the features, dismissing the need for many higher-order terms

in the decision boundary equation that consider the relationship between pairs of dimensions. UMAP is finally applied to non-linearly transform the data to a new distribution, making data of varying gestures more separable. We demonstrate that applying these methods separately can not achieve the desired outcome of the pipeline, with recognition accuracy results lower than the overall pipeline, as shown in Table 2.3. As FDA and PCA alone boost performance for some classifiers, they remain ineffective for some. UMAP can give a consistent recognition performance to all classifiers but at lower accuracy. The pipeline can boost accuracy and consistency for all classifiers while reducing variation between test sessions to a limit. The pipeline features give classifiers a mean accuracy of $86.4\% \pm 8.6$ for session one and $78.5\% \pm 11.0$ for session two, showing a significant enhancement in accuracy and consistency in Figure 2.5. Whereas Figure 2.5b reports a reduced intra-subject error between test sessions. The similar accuracy of classifiers with different complexity degrees indicates that all classifiers learn similar decision boundaries and are subject to the same amount of error. This claim is supported by Figure 2.2b, which depicts data of the same gesture grouped away from other gestures, with a slight overlap for some gestures. This claim is further asserted by Figure 2.6 with different classifiers producing the same misclassifications and predicting the exact wrong gesture for the same input signals.

In addition, We propose a filter-classifier compatibility relationship in Chapter 3 for hand gesture recognition to combat other user-specific inconsistencies. When tested on EMG signals, this relationship is presented as the most considerable influence on gesture recognition performance. We also yield the best processing techniques for EMG gesture recognition when investigated simultaneously, such as varying window size and changing the EMG feature used. Based on the compatibility relationship, a proposed ranking of all investigated classifiers and feature pairs provides future applications and research with an overview of the best compatible features with a classifier, independent of a classifier's complexity. The support vector machine classifier with a linear kernel (SVM-LIN) emphasizes compatibility importance. SVM-LIN is the best classifier pairing for the Willison amplitude feature, at 80.56% recognition accuracy. In contrast, SVM-LIN gives only 69.24% accuracy on the mean absolute value (MAV) feature, despite another linear classifier, linear discriminant analysis (LDA), achieving 84.07%.

LDA performs similarly to the random forest ensemble classifier in Figure 3.9 on the MAV feature. This result ascertains that pairing the proper classifier and feature can result either in the deterioration or enhancement of the recognition performance regardless of the classifier's complexity.

We further conclude that the best filtering window size of the investigated features of EMG in that application is 1250 ms. Increasing the window incorporates more noise, decreasing accuracy, and reducing it does not include enough information to achieve the same accuracy. However, a 250 ms window is recognized as an appropriate filtering window size for real-time applications, below the 300 ms limit, with only a 7% drop in accuracy. The chapter also concludes that (-1, 1) signal normalization, a basic sign-preserving procedure, allows optimal performance. (-1, 1) normalization surpasses raw signals and (0, 1) normalization, a sign-removing normalization.

Our investigation concludes that features with tunable hyperparameters must be tuned separately per user as their optimal value changes from one user to another. This issue can be noticed in Fig. 3.11, for which the threshold values for skewness and kurtosis features were chosen optimally for three users and tested for all 12. However, the recognition results fall lower than any other feature, despite being optimal for three users. Even though the number of bins for the histogram feature was optimized over the same 3 participants and behaved similarly to other features in Figure 3.11, it produces significant outliers. The number of outliers suggests that the mean result of the histogram feature is not a result of consistency for all participants.

We present the methods in this thesis to provide further consistency and stability of electromyography (EMG) and forcemyography (FMG) hand gesture recognition. The findings of this thesis enable lower computation devices to use simple classification models to achieve ensemble classifiers' performance for the hand gesture recognition of electromyography (EMG) and forcemyography (FMG) by tackling several user signal inconsistencies and errors. These methods are also intended to become a basis for future research studies in this field.

Future studies can use our findings as a background for different research purposes. For instance, a future direction might investigate the interpretability of the processed features of our machine learning pipeline as it gives a consistent data distribution regardless of the session.

In addition, our pipeline could be expanded to work with other types of signals for hand gesture recognition. At the same time, the results can be further analyzed to determine if our proposed methods are independent of the participant's biological characteristics. More analytical and processing techniques are available in the literature to expand upon the compatibility pairing ranking we propose in Chapter 3. That chapter's results can also be a basis for investigating other factors that were not included, such as window overlap.

Bibliography

- [1] M. Asfour, C. Menon, and X. Jiang, “A machine learning processing pipeline for reliable hand gesture classification of fmg signals with stochastic variance,” *Sensors*, vol. 21, no. 4, 2021.
- [2] M. Asfour, C. Menon, and X. Jiang, “Feature–classifier pairing compatibility for semg signals in hand gesture recognition under joint effects of processing procedures,” *Bioengineering*, vol. 9, no. 11, 2022.
- [3] A. Dwivedi, Y. Kwon, and M. Liarokapis, “Emg-based decoding of manipulation motions in virtual reality: Towards immersive interfaces,” in *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 09 2020.
- [4] M. Simão, N. Mendes, O. Gibaru, and P. Neto, “A review on electromyography decoding and pattern recognition for human-machine interaction,” *IEEE Access*, vol. PP, pp. 1–1, 03 2019.
- [5] H. Hassan, S. Abou-Loukh, and I. Ibraheem, “Teleoperated robotic arm movement using electromyography signal with wearable myo armband,” *Journal of King Saud University*, vol. 32, pp. 378–387, 05 2019.
- [6] D.-P. Yang, J.-D. Zhao, L. Jiang, and H. Liu, “Embedded online recognition of hand gesture emg modes,” *Harbin Gongye Daxue Xuebao/Journal of Harbin Institute of Technology*, vol. 42, pp. 1060–1065, 07 2010.

- [7] T. Kapuscinski, M. Oszust, M. Wysocki, and D. Warchoł, “Recognition of hand gestures observed by depth cameras,” *International Journal of Advanced Robotic Systems*, vol. 12, 04 2015.
- [8] M. Kim, J. Cho, S. Lee, and Y. Jung, “Imu sensor-based hand gesture recognition for human-machine interfaces,” *Sensors*, vol. 19, p. 3827, 09 2019.
- [9] G. Naik, D. Kumar, V. Singh, and M. Palaniswami, “Hand gestures for hci using ica of emg,” *HCSNet Workshop Use Vis. HCI*, vol. 56, 11 2006.
- [10] N. ha, G. Withanachchi, and Y. Yihun, “Performance of forearm fmg for estimating hand gestures and prosthetic hand control,” *Journal of Bionic Engineering*, vol. 16, pp. 88–98, 01 2019.
- [11] J. Qi, G. Jiang, G. Li, Y. Sun, and B. Tao, “Intelligent human-computer interaction based on surface emg gesture recognition,” *IEEE Access*, vol. PP, pp. 1–1, 05 2019.
- [12] A. Haria, A. Subramanian, N. Asokkumar, S. Poddar, and J. Nayak, “Hand gesture recognition for human computer interaction,” *Procedia Computer Science*, vol. 115, pp. 367–374, 12 2017.
- [13] W. Schweitzer, M. Thali, and D. Egger, “Case-study of a user-driven prosthetic arm design: Bionic hand versus customized body-powered technology in a highly demanding work environment [open access],” *Journal of NeuroEngineering and Rehabilitation*, vol. 15, 01 2018.
- [14] L. Resnik, S. Ekerholm, M. Borgia, and M. A. Clark, “A national study of veterans with major upper limb amputation: Survey methods, participants, and summary findings,” *PLOS ONE*, vol. 14, pp. 1–24, 03 2019.
- [15] U. Pale, M. Atzori, H. Müller, and A. Scano, “Variability of muscle synergies in hand grasps: Analysis of intra-and inter-session data,” *Sensors*, vol. 20, p. 4297, 08 2020.
- [16] M. Zanghieri, S. Benatti, A. Burrello, V. Kartsch, F. Conti, and L. Benini, “Robust real-time embedded emg recognition framework using temporal convolutional networks on a

- multicore iot processor,” *IEEE transactions on biomedical circuits and systems*, vol. PP, 12 2019.
- [17] M. Abdoli-Eramaki, C. Damecour, J. Christenson, and J. Stevenson, “The effect of perspiration on the semg amplitude and power spectrum,” *Journal of electromyography and kinesiology : official journal of the International Society of Electrophysiological Kinesiology*, vol. 22, 05 2012.
- [18] X. Jiang, L. Merhi, and C. Menon, “Force exertion affects grasp classification using force myography,” *IEEE Transactions on Human-Machine Systems*, vol. 48, no. 2, pp. 219–226, 2018.
- [19] Y. Sun, C. Xu, G. Li, W. Xu, J. Kong, D. Jiang, B. Tao, and D. Chen, “Intelligent human computer interaction based on non redundant emg signal,” *Alexandria Engineering Journal*, vol. 59, 01 2020.
- [20] M. Halaki and K. Ginn, “Normalization of emg signals: To normalize or not to normalize and what to normalize to?,” in *Computational Intelligence in Electromyography Analysis* (G. R. Naik, ed.), ch. 7, Rijeka: IntechOpen, 2012.
- [21] A. T. Belyea, K. B. Englehart, and E. J. Scheme, “Fmg versus emg: A comparison of usability for real-time pattern recognition based control,” *IEEE Transactions on Biomedical Engineering*, vol. 66, pp. 3098–3104, 2019.
- [22] X. Jiang, L.-K. Merhi, Z. G. Xiao, and C. Menon, “Exploration of force myography and surface electromyography in hand gesture classification,” *Medical Engineering & Physics*, vol. 41, pp. 63 – 73, 2017.
- [23] Y. Du, W. Jin, W. Wei, Y. Hu, and W. Geng, “Surface emg-based inter-session gesture recognition enhanced by deep domain adaptation,” *Sensors*, vol. 17, no. 3, 2017.
- [24] G. Marano, C. Brambilla, R. M. Mira, A. Scano, H. Müller, and M. Atzori, “Questioning domain adaptation in myoelectric hand prostheses control: An inter- and intra-subject study,” *Sensors*, vol. 21, no. 22, 2021.

- [25] L. H. Smith, L. J. Hargrove, B. A. Lock, and T. A. Kuiken, “Determining the optimal window length for pattern recognition-based myoelectric control: Balancing the competing effects of classification error and controller delay,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 19, no. 2, pp. 186–192, 2011.
- [26] P. Artemiadis and K. Kyriakopoulos, “An emg-based robot control scheme robust to time-varying emg signal,” *Information Technology in Biomedicine, IEEE Transactions on*, vol. 14, pp. 582 – 588, 06 2010.
- [27] A. Prakash, A. Sahi, N. Sharma, and S. Sharma, “Force myography controlled multifunctional hand prosthesis for upper-limb amputees,” *Biomedical Signal Processing and Control*, vol. 62, p. 102122, 09 2020.
- [28] J. Rovira, E. Rocon, D. Reynaerts, B. Saro, S. Levin, and W. Moorleghem, “The manus-hand dextrous robotics upper limb prosthesis: mechanical and manipulation aspects,” *Auton. Robots*, vol. 16, pp. 143–163, 03 2004.
- [29] J. Ribeiro, F. Mota, T. Cavalcante, I. Nogueira, V. Gondim, V. Albuquerque, and A. Alexandria, “Analysis of man-machine interfaces in upper-limb prosthesis: A review,” *Robotics*, vol. 8, p. 16, 02 2019.
- [30] X. Li, O. Samuel, X. Zhang, H. Wang, P. Fang, and P. Li, “A motion-classification strategy based on semg-eeg signal combination for upper-limb amputees,” *Journal of NeuroEngineering and Rehabilitation*, vol. 14, p. 2, 01 2017.
- [31] N. Parajulli, N. Sreenivasan, P. Bifulco, M. Cesarelli, S. Savino, V. Niola, D. Esposito, T. Hamilton, G. Naik, U. Gunawardana, and G. Gargiulo, “Real-time emg based pattern recognition control for hand prostheses: A review on existing methods, challenges and future implementation,” *Sensors*, vol. 19, p. 4596, 10 2019.
- [32] Z. Xiao and C. Menon, “A review of force myography research and development,” *Sensors*, vol. 19, p. 4557, 10 2019.

- [33] S. Jiang, Q. Gao, and H. Liu, "A novel, co-located emg-fmg-sensing wearable armband for hand gesture recognition," *Sensors and Actuators A Physical*, vol. 301, p. 111738, 01 2020.
- [34] A. Radmand, E. Scheme, and K. Englehart, "High-density force myography: A possible alternative for upper-limb prosthetic control," *Journal of Rehabilitation Research and Development*, vol. 53, pp. 443–456, 07 2016.
- [35] A. Belyea, K. Englehart, and E. Scheme, "Fmg vs emg: A comparison of usability for real-time pattern recognition based control," *IEEE Transactions on Biomedical Engineering*, vol. PP, pp. 1–1, 02 2019.
- [36] D. Tkach, H. Huang, and T. Kuiken, "Study of stability of time-domain features for electromyographic pattern recognition," *Journal of neuroengineering and rehabilitation*, vol. 7, p. 21, 05 2010.
- [37] I. Ketykó, F. Kovács, and K. Z. Varga, "Domain adaptation for semg-based gesture recognition with recurrent neural networks," *2019 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–7, 2019.
- [38] Y. Du, W. Jin, W. Wei, Y. Hu, and W. Geng, "Surface emg-based inter-session gesture recognition enhanced by deep domain adaptation," *Sensors*, vol. 17, p. 458, 02 2017.
- [39] B. V. M. Patil, "Pca and fda based dimensionality reduction techniques for effective fault diagnosis of rolling element bearing," *Journal of emerging technologies and innovative research*, vol. 2, p. 1297, 2015.
- [40] I. Jolliffe and J. Cadima, "Principal component analysis: A review and recent developments," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 374, p. 20150202, 04 2016.
- [41] L. McInnes, J. Healy, N. Saul, and L. Grossberger, "Umap: Uniform manifold approximation and projection," *Journal of Open Source Software*, vol. 3, p. 861, 09 2018.

- [42] L. Wander, A. Vianello, J. Vollertsen, F. Westad, U. Braun, and A. Paul, “Exploratory analysis of hyperspectral ftir data obtained from environmental microplastics samples,” *Analytical Methods*, vol. 12, 01 2020.
- [43] A. Diaz-Papkovich, L. Anderson-Trocmé, C. Ben-Eghan, and S. Gravel, “Umap reveals cryptic population structure and phenotype heterogeneity in large genomic cohorts,” *PLOS Genetics*, vol. 15, p. e1008432, 11 2019.
- [44] P. Xanthopoulos, P. Pardalos, and T. Trafalis, “Linear discriminant analysis,” *Robust Data Mining*, pp. 27–33, 01 2013.
- [45] A. Godiyal, M. Mondal, S. Joshi, and D. Joshi, “Force myography based novel strategy for locomotion classification,” *IEEE Transactions on Human-Machine Systems*, vol. PP, pp. 1–10, 08 2018.
- [46] C. Ahmadizadeh, L. Merhi, B. Pousett, S. Sangha, and C. Menon, “Toward intuitive prosthetic control: Solving common issues using force myography, surface electromyography, and pattern recognition in a pilot case study,” *IEEE Robotics Automation Magazine*, vol. 24, no. 4, pp. 102–111, 2017.
- [47] G. Sadarangani, X. Jiang, L. Simpson, J. Eng, and C. Menon, “Force myography for monitoring grasping in individuals with stroke with mild to moderate upper-extremity impairments: A preliminary investigation in a controlled environment,” *Frontiers in Bioengineering and Biotechnology, section Bionics and Biomimetics*, vol. 5, 07 2017.
- [48] A. Tharwat, “Linear vs. quadratic discriminant analysis classifier: a tutorial,” *International Journal of Applied Pattern Recognition*, vol. 3, p. 145, 01 2016.
- [49] N. Guenther and M. Schonlau, “Support vector machines,” *Stata Journal*, vol. 16, pp. 917–937, 01 2016.
- [50] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, pp. 436–44, 05 2015.
- [51] L. Peterson, “K-nearest neighbor,” *Scholarpedia*, vol. 4, p. 1883, 01 2009.

- [52] G. Y, N. Nair, P. Satpathy, and J. Christopher, “Covariate shift: A review and analysis on classifiers,” in *2019 Global Conference for Advancement in Technology (GCAT)*, pp. 1–6, 10 2019.
- [53] A. Dwivedi, Y. Kwon, and M. Liarokapis, “Emg-based decoding of manipulation motions in virtual reality: Towards immersive interfaces,” in *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 09 2020.
- [54] H. Hassan, S. Abou-Loukh, and I. Ibraheem, “Teleoperated robotic arm movement using electromyography signal with wearable myo armband,” *Journal of King Saud University*, vol. 32, pp. 378–387, 05 2019.
- [55] H. Liu, D. Yang, L. Jiang, and S. Fan, “Development of a multi-dof prosthetic hand with intrinsic actuation, intuitive control and sensory feedback,” *Industrial Robot*, vol. 41, pp. 381–392, 08 2014.
- [56] T. Kapuscinski, M. Oszust, M. Wysocki, and D. Warchoł, “Recognition of hand gestures observed by depth cameras,” *International Journal of Advanced Robotic Systems*, vol. 12, 04 2015.
- [57] M. Kim, J. Cho, S. Lee, and Y. Jung, “Imu sensor-based hand gesture recognition for human-machine interfaces,” *Sensors*, vol. 19, p. 3827, 09 2019.
- [58] G. Naik, D. Kumar, V. Singh, and M. Palaniswami, “Hand gestures for hci using ica of emg,” *HCSNet Workshop Use Vis. HCI*, vol. 56, 11 2006.
- [59] D.-P. Yang, J.-D. Zhao, L. Jiang, and H. Liu, “Embedded online recognition of hand gesture emg modes,” *Harbin Gongye Daxue Xuebao/Journal of Harbin Institute of Technology*, vol. 42, pp. 1060–1065, 07 2010.
- [60] M. Simão, N. Mendes, O. Gibaru, and P. Neto, “A review on electromyography decoding and pattern recognition for human-machine interaction,” *IEEE Access*, vol. 7, pp. 39564–39582, 2019.

- [61] H. F. Hassan, S. J. Abou-Loukh, and I. K. Ibraheem, "Teleoperated robotic arm movement using electromyography signal with wearable myo armband," *Journal of King Saud University - Engineering Sciences*, vol. 32, no. 6, pp. 378–387, 2020.
- [62] J. Ribeiro, F. Mota, T. Cavalcante, I. Nogueira, V. Gondim, V. Albuquerque, and A. Alexandria, "Analysis of man-machine interfaces in upper-limb prosthesis: A review," *Robotics*, vol. 8, no. 1, 2019.
- [63] X. Li, O. Samuel, X. Zhang, H. Wang, P. Fang, and P. Li, "A motion-classification strategy based on semg-eeeg signal combination for upper-limb amputees," *Journal of NeuroEngineering and Rehabilitation*, vol. 14, p. 2, 01 2017.
- [64] A. Prakash, S. Sharma, and N. Sharma, "A compact-sized surface emg sensor for myoelectric hand prosthesis," *Biomedical Engineering Letters*, vol. 9, 08 2019.
- [65] A. Dwivedi, Y. Kwon, and M. Liarokapis, "Emg-based decoding of manipulation motions in virtual reality: Towards immersive interfaces," 09 2020.
- [66] N. Parajulli, N. Sreenivasan, P. Bifulco, M. Cesarelli, S. Savino, V. Niola, D. Esposito, T. Hamilton, G. Naik, U. Gunawardana, and G. Gargiulo, "Real-time emg based pattern recognition control for hand prostheses: A review on existing methods, challenges and future implementation," *Sensors*, vol. 19, p. 4596, 10 2019.
- [67] F. AlOmari and G. Liu, "Analysis of extracted forearm semg signal using lda, qda, k-nn classification algorithms," *The Open Automation and Control Systems Journal*, vol. 6, pp. 108–116, 07 2014.
- [68] C. Spiewak, M. R. Islam, M. Assad-Uz-Zaman, and M. Rahman, "A comprehensive study on emg feature extraction and classifiers," *Open Access Journal of Biomedical Engineering and its Applications*, vol. 1, 02 2018.
- [69] D. Tkach, H. Huang, and T. Kuiken, "Study of stability of time-domain features for electromyographic pattern recognition," *Journal of neuroengineering and rehabilitation*, vol. 7, p. 21, 05 2010.

- [70] A. Phinyomark, P. Phukpattaranont, and C. Limsakul, "Feature reduction and selection for emg signal classification," *Expert Systems with Applications*, vol. 39, p. 7420–7431, 06 2012.
- [71] H. Ling, Y. Bo, and Z. Lina, "Clustering analysis and recognition of the emgs," 07 2011.
- [72] H. Mizuno, N. Tsujiuchi, and T. Koizumi, "Forearm motion discrimination technique using real-time emg signals," *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*, vol. 2011, pp. 4435–8, 08 2011.
- [73] C. Savur and F. Sahin, "Real-time american sign language recognition system using surface emg signal," pp. 497–502, 12 2015.
- [74] M. E. Benalcázar, A. G. Jaramillo, Jonathan, A. Zea, A. Páez, and V. H. Andaluz, "Hand gesture recognition using machine learning and the myo armband," in *2017 25th European Signal Processing Conference (EUSIPCO)*, pp. 1040–1044, 2017.
- [75] J. Mendes Júnior, M. Freitas, H. Siqueira, A. Lazzaretti, S. Pichorim, and S. Stevan Jr, "Feature selection and dimensionality reduction: An extensive comparison in hand gesture classification by semg in eight channels armband approach," *Biomedical Signal Processing and Control*, vol. 59, p. 101920, 05 2020.
- [76] R. Ekstrom, G. Soderberg, and R. Donatelli, "Normalization procedures using maximum voluntary isometric contractions for serratus anterior and trapezius muscles during surface emg analysis," *Journal of electromyography and kinesiology : official journal of the International Society of Electrophysiological Kinesiology*, vol. 15, pp. 418–28, 09 2005.
- [77] A. Morris, G. Kemp, A. Lees, and S. Frostick, "A study of reproducibility of three different normalization methods in intramuscular dual fine wire electromyography of the shoulder," *Journal of electromyography and kinesiology : official journal of the International Society of Electrophysiological Kinesiology*, vol. 8, pp. 317–22, 11 1998.

- [78] D. Rouffet and C. Hautier, “Emg normalization to study muscle activation in cycling,” *Journal of electromyography and kinesiology : official journal of the International Society of Electrophysiological Kinesiology*, vol. 18, pp. 866–78, 05 2007.
- [79] X. Jiang, L.-K. Merhi, Z. Xiao, and C. Menon, “Exploration of force myography and surface electromyography in hand gesture classification,” *Medical Engineering & Physics*, vol. 41, p. In Press, 02 2017.
- [80] M. Cutkosky, “On grasp choice, grasp models, and the design of hands for manufacturing tasks,” *IEEE Transactions on Robotics and Automation*, vol. 5, no. 3, pp. 269–279, 1989.
- [81] J.-W. Lin, C. Wang, Y. Y. Huang, K.-T. Chou, H.-Y. Chen, W.-L. Tseng, and M. Y. Chen, “Backhand: Sensing hand gestures via back of the hand,” in *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, UIST ’15, (New York, NY, USA), p. 557–564, Association for Computing Machinery, 2015.
- [82] T. M. Skirven, A. L. Osterman, J. Fedorczyk, and P. C. Amadio, *Rehabilitation of the hand and upper extremity, 2-volume set E-book: expert consult*. Elsevier Health Sciences, 2011.
- [83] K. Kumari and S. Yadav, “Linear regression analysis study,” *Journal of the Practice of Cardiovascular Sciences*, vol. 4, p. 33, 01 2018.
- [84] P. Xanthopoulos, P. Pardalos, and T. Trafalis, “Linear discriminant analysis,” *Robust Data Mining*, pp. 27–33, 01 2013.
- [85] N. Guenther and M. Schonlau, “Support vector machines,” *Stata Journal*, vol. 16, pp. 917–937, 01 2016.
- [86] A. Tharwat, “Linear vs. quadratic discriminant analysis classifier: a tutorial,” *International Journal of Applied Pattern Recognition*, vol. 3, p. 145, 01 2016.
- [87] I. Rish, “An empirical study of the naïve bayes classifier,” *IJCAI 2001 Work Empir Methods Artif Intell*, vol. 3, 01 2001.

- [88] J. R. Quinlan, "Induction of decision trees," *Machine Learning*, vol. 1, pp. 81–106, 2004.
- [89] L. Peterson, "K-nearest neighbor," *Scholarpedia*, vol. 4, p. 1883, 01 2009.
- [90] L. Breiman, "Random forests," *Machine Learning*, vol. 45, pp. 5–32, 10 2001.
- [91] A. Natekin and A. Knoll, "Gradient boosting machines, a tutorial," *Frontiers in neuro-robotics*, vol. 7, p. 21, 12 2013.
- [92] C. Luca, "The use of surface electromyography in biomechanics," *Journal of Applied Biomechanics*, vol. 13, pp. 135–163, 05 1997.
- [93] P. Konrad, "The abc of emg," *A practical introduction to kinesiological electromyography*, vol. 1, no. 2005, pp. 30–5, 2005.

Appendix A

Interdisciplinary Committee on Ethics in Human Research (ICEHR) Approval



**Interdisciplinary Committee on
Ethics in Human Research (ICEHR)**

St. John's, NL Canada A1C 5S7
Tel: 709 864-2561 icehr@mun.ca
www.mun.ca/research/ethics/humans/icehr

ICEHR Number:	20230479-SC
Approval Period:	July 7, 2022 – July 31, 2023
Funding Source:	
Responsible Faculty:	Dr. Xianta Jiang Department of Computer Science
Title of Project:	<i>Machine Learning and Processing Techniques for the Enhancement of Hand Gesture Recognition of Forcemyography and Electromyography Signals</i>

July 7, 2022

Asfour Mohammed
Department of Computer Science, Faculty of Science
Memorial University

Dear Asfour Mohammed:

Thank you for your submission to the Interdisciplinary Committee on Ethics in Human Research (ICEHR), seeking ethical clearance for your research project. The Committee appreciates the care and diligence with which you prepared your application. The project is consistent with the guidelines of the *Tri-Council Policy Statement on Ethical Conduct for Research Involving Humans (TCPS2)*. *Full ethics clearance* is granted for **one year** from the date of this letter. ICEHR approval applies to the ethical acceptability of the research, as per Article 6.3 of the *TCPS2 (2018)*. Researchers are responsible for adherence to any other relevant University policies and/or funded or non-funded agreements that may be associated with the project. If funding is obtained subsequent to ethics approval, you must submit a Funding and/or Partner Change Request to ICEHR so that this ethics clearance can be linked to your award.

The *TCPS2* **requires** that you **strictly adhere to the protocol and documents as last reviewed** by ICEHR. If you need to make additions and/or modifications, you must submit an Amendment Request with a description of these changes, for the Committee's review of potential ethical issues, before they may be implemented. Submit a Personnel Change Form to add or remove project team members and/or research staff. Also, to inform ICEHR of any unanticipated occurrences, an Adverse Event Report must be submitted with an indication of how the unexpected event may affect the continuation of the project.

The *TCPS2* **requires** that you submit an Annual Update to ICEHR before **July 31, 2023**. If you plan to continue the project, you need to request renewal of your ethics clearance and include a brief summary on the progress of your research. When the project no longer involves contact with human participants, is completed and/or terminated, you are required to provide an annual update with a brief final summary and your file will be closed. All post-approval ICEHR event forms noted above must be submitted by selecting the **Applications: Post-Review** link on your Researcher Portal homepage. We wish you success with your research.

Yours sincerely,

Kelly Blidook, Ph.D.
Chair, Interdisciplinary Committee on
Ethics in Human Research

KB/bc

copy: Supervisor – Dr. Xianta Jiang, Department of Computer Science