An Exploration of the Optimal Feature-Classifier Combinations for Transradial Prosthesis Control

Fraser Douglas[1], *Graduate Student Member, IEEE*, Cheryl Docherty, Harry Gover,
Gordon Shields, Konstantina Leventi, Gaetano Di Caterina, *Member, IEEE*

***Abstract*—Within state-of-the-art gesture-based upper-limb myoelectric prosthesis control, gesture recognition commonly relies on the classification of features extracted from electromygraphic (EMG) data gathered from the amputee's residual forearm musculature. Despite best efforts in broadly maximizing gesture recognition accuracy, there does not yet exist a feature-classifier combination accepted as best-practice. In turn, this work hypothesizes that no single feature-classifier combination can consistently maximize accuracy across subjects, positing instead that control schemes should be personalized to the individual. To investigate this hypothesis, the study employed the 40-subject, 49-gesture Ninapro DB2 to compare the performance of 7 different historic, more recent and state-of-the-art feature sets, in combination with 5 machine learning classifiers commonly seen within EMG-based pattern recognition literature. The results demonstrate the ability of Linear Discriminant Analysis (LDA) to marginally exceed other more computationally intensive classifiers in terms of mean accuracy, while the feature set which maximized the highest proportion of individuals' accuracies was shown to vary with both classifier choice and gesture count.**

I. Introduction

Commercial upper-limb myoelectric prostheses have afforded amputees a level of independence not previously attainable. Despite these benefits, the intuitiveness of such devices is often restricted as they rely on simplistic amplitude-based on-off control schemes, detecting simply open or close gestures. Although the users' degrees-of-freedom may be expanded through the use of state-machines or accompanying smartphone applications, this has the potential to lead to device abandonment as the actions performed by the user are not directly mimicked by their prosthesis. It is through pattern-recognition that one might address these issues, increasing the degrees-of-freedom available to user while ensuring control remains intuitive, with lab-based investigations showing the ability to accurately classify as many as 52 unique hand gestures [1], [2]. Despite the success of these studies, there is not yet an agreed upon best practice when classifying these highly stochastic, multi-channel electromyographic (EMG) signals.

Traditional EMG pattern-recognition (EMG-PR) generally follows a set processing sequence; the signal is first pre-processed to remove undesirable frequencies (motion artefacts, mains interference, high frequency noise etc.), then windowed as instantaneous EMG amplitudes present little information about the action performed by the user. At this stage, feature extraction occurs, allowing the data corresponding to different hand gestures to be more easily differentiated during classification [3], [4]. It is the selection of these features which has been shown to have the largest bearing on classification accuracy [5]. As technology has advanced, more complex, higher dimensional feature sets have been explored, with time and time-frequency domain features showing the greatest success [6]. Despite this, there still exist a myriad potential feature combinations that can enhance prosthesis control.

When comparing the accuracies achieved by similar feature sets across studies, performance can vary significantly with data acquisition, pre-processing and windowing protocols being inconsistent. Despite this, computationally low time domain feature sets, including that proposed by Phinyomark et al. [6], have shown the ability to outperform similar notable time domain sets, including the classic Hudgins features which represent the closest thing to a standard set within EMG-PR [9], [12]. Not included within this comparison were more complex features such as the EMG histogram and the notable time-frequency wavelet features which have been implemented to varying degrees of success [11], [13]. The inclusion of these within the feature vector have been shown to improve performance beyond that of more traditional feature sets, albeit only marginally [2], [11]. There also exist more novel feature sets such as the Temporal-Spatial Descriptors (TSD) which have seen limited use within literature yet have shown the potential to outperform all sets mentioned prior, albeit when used with a reduced number of gestures and a limited subject pool [14].

Aside from the choice of feature set, another important consideration is the classification algorithm that interprets the feature vector to identify the gesture performed by the user. Despite the apparent importance of this consideration, D. Tkach has stated that 'the type of classifier used does not significantly affect the classification performance' [8]. While this has been shown to be true in part, the widely used, computationally low Linear Discriminant Analysis (LDA) has shown potential in outperforming other common machine learning algorithms, including Support Vector Machines (SVM) and k Nearest Neighbors (kNN), when classifying 5 gestures using solely time domain features [9]. In a similar study, incorporating an increased number of features within both the time and time-frequency domain to classify 4 unique hand gestures, LDA again showed the ability to outperform more complex linear and non-linear artificial neural networks [10]. In contrast, alternative studies employing increased gesture counts have concluded that LDA can be outperformed by SVM, kNN as well as Random Forests (RF) which maximised the mean classification accuracy [2] even when compared with state-of-the-art Convolutional Neural Networks (CNNs) [11] that have, on occasion, outperformed more traditional classification schemes [1].

Despite the successes reported when using dimensionally heavy feature sets and computationally intensive classifiers, there still remains a large gap between what is technically feasible and what is fit for implementation. It is for this reason that work must continue to explore classical, computationally low EMG-PR techniques with a look to understanding which feature set, given the classifier, individual and the number of gestures they seek to employ, will maximize their gesture recognition accuracy. With this in mind, in selecting the feature sets and classification algorithms for this investigation, it was important to balance novelty and historical relevance. Furthermore, this paper will seek to not only investigate how the mean classification accuracy varies across classifiers and feature sets, but also which feature sets for each classifier provide the highest accuracy for the greatest majority of subjects, as well as how these results change as the number of gestures varies.

II. Methodology

This study explored the performance of 7 different feature sets classified by 5 common machine learning algorithms. These feature sets, described in detail within section II.A., were the Hudgins, Phinyomark, Wavelet Packet Transform (WPT), Histogram (Hist), Phinyo-Hist (P-H), Time Domain Descriptor (TDD) and TSD sets. The chosen classification algorithms were LDA, Quadratic Discriminant Analysis (QDA), RF, SVM and kNN. All investigations were conducted using MATLAB R2020b with EMG data sourced from the publicly available Ninapro DB2 [2], [7]. The creation of this dataset is described within the paper by M. Atzori et al. [7] and features 6 consecutive repetitions of 49 unique hand gestures performed by 40 intact subjects. These signals were windowed at 200 ms with a 75 ms overlap. [2],

In generating the results of Section III., all feature-classifier combinations were used for the full 49 gestures. To then allow for the reduction of this gesture count, the top performing 'n' gestures in terms of their mean true positive rate across all subjects for each specific feature-classifier combination were chosen to represent the gesture subset to be subsequently classified. From here, each feature-classifier combination's performance was assessed for their specific top performing 8, 16, 24, 32 and 40 gestures.

A. Feature Selection & Extraction

To allow for a suitable comparison in terms of novel feature set performance, the low-complexity standard time domain set proposed by Hudgins [12], alongside the similarly low-complexity Phinyomark (or Phin) set, were selected for inclusion within the study. The Hudgins set is comprised of the mean absolute value (MAV), zero crossings (ZC), slope sign changes (SSC) and waveform length (WL), with Phinyomark's differing in that, following an investigation into individual feature redundancy [6], ZC is replaced by the Wilson amplitude (WAMP) and 4^{th} order autoregressive coefficient (AR4). Hudgins extraction was performed with reference to the equations presented by D. Tkach [8], with the additional features (AR4, WAMP) required by Phinyomark's set, extracted in line with the paper by A. Phinyomark [6].

The WPT, used by Englehart et al. [15], has been shown to outperform other time-frequency domain features within EMG-PR. This set was computed using a db45 wavelet at a decomposition level of 3 with statistical features then extracted from each of the 8 end nodes to form the final WPT feature set. These features were MAV, WL, variance, sum of absolute values, simple square integral, standard deviation, RMS natural logarithm and base 10 RMS logarithm.

The EMG histogram, representing the signal amplitude distribution across the temporal window [11], has shown the greatest success when paired with other time domain features [8]. As such, this study paired the histogram with the Hudgins and Phinyomark sets to form the Histogram (Hist) and Phinyo-Hist (P-H) sets. For implementation, the EMG histogram distributed the signal amplitudes within each window across 20 bins in line with literature [2].

Beyond these, the TSD set proposed by Khushaba et al. [16] involves the extraction of TDD features estimating the signal's power spectrum. TDD features are extracted from each EMG channel, as well as the differences between each channel, representing how muscles' relationships change over time [14]. It is the concatenation of the correlation coefficients computed from the TDD features which forms the TSD set. These were computed with reference to the methods outlined by Khushaba et al. [16].

B. Classifier Implementation

The five machine learning classifiers included within this study represented those seen most commonly within EMG-PR literature. For LDA and QDA, as well as SVM using the default linear kernel, the standard Mathworks implementation was followed. For RF, a uniform class probability was assumed and a forest size of 50 was chosen to allow for suitably high accuracies while ensuring computational complexity and the probability of overfitting remained low. For kNN, various distance functions were assessed in terms of their mean classification accuracy before settling on the city block distance function with a cluster count equal to the number of classes. To confirm the accuracy of the results, a technique akin to leave-one-out cross-validation was employed. Here, 2 non-adjacent repetitions, out of the 6 sequential repetitions of each unique gesture, were placed into the testing set and the remaining 4 into the training set [2]. This allowed 10 cross-validation iterations, ensuring data diversity and improving confidence in results.

C. Dimensionality Reduction

When using higher dimensionality feature sets, to ensure that long processing times are not incurred, a dimensionality reduction technique is often implemented, with the most common within EMG-PR being Principal Component Analysis (PCA) [17], [18]. With PCA, the data is projected onto orthogonal vectors within a new coordinate frame using eigenvector decomposition where successive components maximize variance, allowing one to explain the majority of the data's variance in a reduced number of dimensions [19]. One alternative technique is Spectral Regression (SR) [20], [21], whose potential has been shown by Khushaba et al. when applied to the TSD set [16]. In contrast to PCA, SR's computation does not involve the eigen-decomposition of the feature matrix, instead decomposing the subspace learning into two-steps: graph embedding for response learning and regression for projective function learning [20]. This reduces the computational complexity over PCA, making it better suited for real-time control.

In deciding the dimensionality reduction method for each feature set, subjects 1-5 were taken as a representative subgroup with all 49 gestures investigated. In each case, all feature-classifier combinations were explored following SR, PCA and, aside from when computation was prohibitively expensive, no dimensionality reduction. Results showed that, for the low dimensionality Hudgins and Phinyomark's sets, no dimensionality reduction maximized performance. In all other cases, SR outperformed PCA both in terms of accuracy and runtime. In response, SR was used for all feature sets with the exception of Hudgins and Phinyomark's, where no dimensionality reduction was used, and WPT, where PCA was employed as this is the standard method used with WPT within literature.

III. RESULTS

The full results from this study are presented within Figure 1 where it can be seen that, for the full 49 gestures, the combination of kNN and TDD produced the highest mean classification accuracy of 66.0%, closely followed by RF and Phinyomark at 65.6%. This was the sole gesture count investigated where LDA did not achieve the maximum classification accuracy, as indicated by Figure 1b. At 8 gestures, LDA and P-H obtained the highest mean accuracy of 85.9% but, despite this, only 14% of subjects achieved their highest accuracy with P-H compared to 56% maximizing their accuracy with Phinyomark's set. From Figure 1c, it is clear how the utility of select feature sets can vary dramatically with gesture count, with Phinyomark's set accounting for 97% of subjects' maxima at 49 gestures with RF reducing to just 11% with 8 gestures despite its consistently high classification accuracy when compared to the other feature sets.

Beyond this, the results indicate how feature set performance can vary drastically across classifiers, where the mean accuracies of the TDD and TSD sets decreased from 66.0% and 64.9% with kNN, to 6.3% and 6.5% with SVM, when using the full 49 gestures. A similar effect, albeit to a lesser extent, was seen with WPT whereby the accuracies achieved by LDA and QDA remained similar regardless of gesture count, while a marked reduction in accuracy was seen when used with RF, SVM or kNN, with kNN causing the greatest decrease in accuracy.

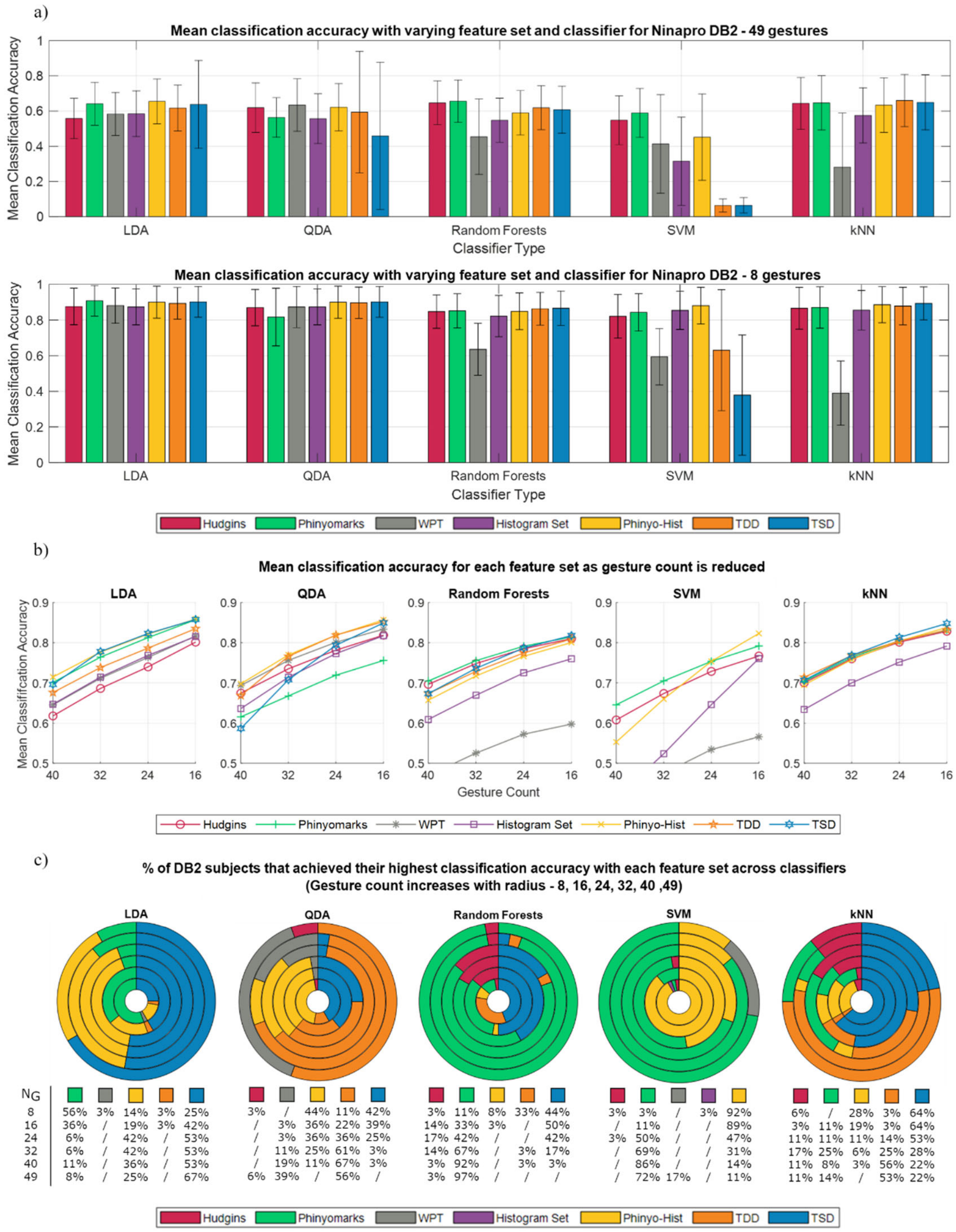

Figure 1. Results from the feature-classifier investigation as gesture count varies. a), mean classification accuracy across all subjects for 49 and 8 gestures with standard deviation error bars. b), mean classification accuracies across subjects for each feature set as gesture count is reduced from 40 to 16. c), % of subjects that presented their highest accuracy with each feature set across all gesture counts.

IV. Discussion

This study acted to explore how the optimal feature-classifier combinations used for gesture recognition varied with gesture count. In general, the results act to partially support D. Tkach's claim that 'the type of classifier used does not significantly affect classification performance' [8]. This held true for Hudgins and Phinyomarks sets, while the results indicate that certain classifiers are wholly unsuitable for select feature sets. As such, to propose a revision to the original statement, for select feature sets, classifier choice affects performance minimally, while for others the incorrect choice of classifier has the potential to dramatically reduce classification accuracy.

When evaluating general performance, LDA was seen to have the highest mean classification accuracy across all combined feature sets. This, in combination with it providing the highest mean accuracy for each gesture count other than 49, well exemplifies its usefulness within EMG-PR, making it the clear choice for evaluating the general performance of novel feature sets regardless of the gesture count. This being said, more online control studies are needed before being able to say conclusively that LDA represents the optimal choice of classifier for real time prosthesis control.

Looking broadly at the results, it is clear how the overall mean classification accuracy can be misleading if used with the goal of predicting which feature-classifier combination will maximize a subject's accuracy. As an example, if one has decided to use SVM to classify 49 gestures and used solely the results of Figure 1a to drive their feature set choice, WPT would likely be discounted due to its low accuracy which was greatly exceeded by the Hudgins, Phinyomark and P-H sets. Despite this, 17% of all subjects obtained their highest accuracy with WPT, while 11% and 0% obtained their maximum accuracy with P-H and Hudgins respectively. This is likely due to the fact that, in terms of the information they encode, there is a large overlap between Hudgins, Phinyomarks, the Histogram and P-H sets.

This, in turn, exemplifies the need for a more personalized approach when constructing an individual's control scheme as, even if one feature set performs poorly in general, if the information it encodes is unique then there is potential for it to outperform other more robust feature sets that showed a higher mean classification accuracy. The hope is that these results can help guide classifier and feature selection for future studies, with the goal of understanding how these results translate to real time performance such that the intuitiveness of control in upper-limb myoelectric prostheses might be improved and the number of degrees of freedom available to the end-user are maximized.

V. Conclusion

In conclusion, this study acted to confirm the idea that no single feature-classifier combination explored here can consistently maximize gesture recognition accuracy across subjects, exemplifying the need for a more individualistic approach when considering amputee usability. This being said, the power of LDA has again been shown, offering consistently high classification accuracies for feature sets where other classifiers struggled, making it the clear choice in the assessment of novel feature sets. Moving forward, increased resources should be spent exploring the best methods of maximizing individual online accuracies while ensuring computational complexity remains low. This should in turn help in promoting further research focused on the clinical translation of EMG-PR prosthesis control, a commercially underutilized technology with huge potential in enriching the daily lives of upper-limb amputees.

VI. References

[1] W. Geng, Y. Du, W. Jin, W. Wei, Y. Hu, and J. Li, "Gesture recognition by instantaneous surface EMG images," *Sci. Rep. 2016 6:1*, vol. 6, no. 1, pp. 1–8, Nov. 2016

[2] M. Atzori *et al.*, "Electromyography data for non-invasive naturally-controlled robotic hand prostheses," *Sci. Data 2014 1:1*, vol. 1, no. 1, pp. 1–13, Dec. 2014

[3] K. Englehart and B. Hudgins, "A Robust, Real-Time Control Scheme for Multifunction Myoelectric Control," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 7, pp. 848–854, 2003

[4] M. Asghari Oskoei and H. Hu, "Myoelectric control systems—A survey," *Biomed. Signal Process. Control*, vol. 2, no. 4, Oct. 2007

[5] P. Geethanjali, "Myoelectric control of prosthetic hands: state-of-the-art review," *Med. Devices (Auckl).*, vol. 9, pp. 247–255, Jul. 2016

[6] A. Phinyomark, P. Phukpattaranont, and C. Limsakul, "Feature reduction and selection for EMG signal classification," *Expert Syst. Appl.*, vol. 39, no. 8, pp. 7420–7431, Jun. 2012

[7] M. Atzori *et al.*, "Building the Ninapro database: A resource for the biorobotics community," *Proceedings of the IEEE BioRob*, pp. 1258–1265, 2012

[8] D. Tkach, H. Huang, and T. A. Kuiken, "Study of stability of time-domain features for electromyographic pattern recognition," *J. NeuroEng. Rehabil.*, vol. 7, no. 1, pp. 1–13, May 2010

[9] A. Radmand, E. Scheme, P. Kyberd, and K. Englehart, "Investigation of Optimum Pattern Recognition Methods for Robust Myoelectric Control During Dynamic Limb Movement," *CMBES Proceedings*, vol. 36, May 2013

[10] A. A. Adewuyi, L. J. Hargrove, and T. A. Kuiken, "Evaluating eMg Feature and classifier selection for application to Partial-hand Prosthesis control," *Front. Neurorob.*, vol. 10, p. 15, 2016

[11] M. Atzori, M. Cognolato, and H. Müller, "Deep Learning with Convolutional Neural Networks Applied to Electromyography Data: A Resource for the Classification of Movements for Prosthetic Hands," *Front. Neurorob.*, vol. 10, Sep. 2016

[12] B. Hudgins, P. Parker, and R. N. Scott, "A New Strategy for Multifunction Myoelectric Control," *IEEE Trans. Biomed. Eng.*, vol. 40, no. 1, pp. 82–94, 1993

[13] M. Atzori and M. Henning, "PaWFE: Fast Signal Feature Extraction Using Parallel Time Windows," *Front. Neurorob.*, vol. 13, p. 74, 2019

[14] R. N. Khushaba, A. H. Al-Timemy, A. Al-Ani, and A. Al-Jumaily, "A Framework of Temporal-Spatial Descriptors-Based Feature Extraction for Improved Myoelectric Pattern Recognition," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 10, pp. 1821–1831, Oct. 2017

[15] K. Englehart, B. Hudgins, and P. A. Parker, "A wavelet-based continuous classification scheme for multifunction myoelectric control," *IEEE Trans. Biomed. Eng.*, vol. 48, no. 3, pp. 302–311, 2001

[16] R. N. Khushaba, A. Al-Timemy, A. Al-Ani, and A. Al-Jumaily, "Myoelectric feature extraction using temporal-spatial descriptors for multifunction prosthetic hand control," *Proceedings of the Annual International Conference of the IEEE EMBS*, pp. 1696–1699, Oct. 2016

[17] M. B. I. Reaz, M. S. Hussain, and F. Mohd-Yasin, "Techniques of EMG signal analysis: detection, processing, classification and applications," *Biological Procedures Online*, vol. 8, no. 1, p. 11, Mar. 2006,

[18] P. Geethanjali, "Comparative study of PCA in classification of multichannel EMG signals," *Australasian Physical and Engineering Sciences in Medicine*, vol. 38, no. 2, pp. 331–343, Jun. 2015

[19] I. T. Jollife and J. Cadima, "Principal component analysis: a review and recent developments," *Philos. Trans. R. Soc. London, Ser. A*, vol. 374, no. 2065, Apr. 2016

[20] D. Cai, X. He, and J. Han, "Spectral regression for efficient regularized subspace learning," *Proceedings of the IEEE International Conference on Computer Vision*, 2007

[21] D. Cai, X. He, and J. Han, "SRDA: An efficient algorithm for large scale discriminant analysis," *IEEE Trans. Knowl. Data Eng.*, vol. 20, no. 1, pp. 1–12, 2008

[22] A. Jaramillo-Yánez, M. E. Benalcázar, and E. Mena-Maldonado, "Real-Time Hand Gesture Recognition Using Surface Electromyography and Machine Learning: A Systematic Literature Review," *Sensors*, vol. 20, no. 9, p. 2467, Apr. 2020