ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

EMG Signal Analysis

Mini Project 2 Report

NX421 - Neural Signals and Signal Processing

Group Members:

Angana Mondal
Heliya Shakeri
Khushi Singh
Arnault Dominic Philippe Stähli
Pamela Van Den Enden Uribe

16th January 2026

1 Part One: Single subject classification

1.1 Pre-processing and Feature extraction

In this part of the project, the goal is to classify 12 movement classes from surface EMG signals recorded from a single subject in the NinaPro Database Atzori, Gijsberts, Kuzborskij et al. 2015. Each movement was repeated 10 times, with signals acquired from 10 channels at 100 Hz. Upon inspecting the dataset, we found that it had been preprocessed: the signals are rectified, there is no noticeable power-line noise, and the frequency range appears reasonable, so no bandpass filtering was necessary. Visualization of each channel confirmed signal consistency across repetitions, with no constant or outlier channels, so no data was excluded.

Feature extraction. Feature selection was based on Atzori, Gijsberts, Castellini et al. 2014. The time-domain features used in such article included Root Mean Square (RMS), Mean Absolute Value (MAV), Waveform Length (WL), Slope Sign Changes (SSC), and Willison Amplitude (WAMP), along with histogram-based features and marginal Discrete Wavelet Transform (mDWT) features. Variance (VAR) and Integrated EMG (IEMG) were added based on lecture material, and Power Spectral Density (PSD) was included to incorporate a frequency-domain component. By combining these complementary features, the extracted set provides information on signal amplitude, shape, complexity, temporal patterns, and frequency content, thereby enhancing the ability to distinguish among different stimuli from EMG recordings.

Figure 1.1 shows that the RMS, MAV, and WL features vary somewhat across channels while remaining relatively consistent within each channel across repetitions of the same movement, despite some variation between different movements. Other features show similar patterns. This can be explained by the fact that each channel captures activity from a specific muscle group, which produces distinct signal features, while repeated executions of the same movement generate similar muscle activation patterns within that channel.

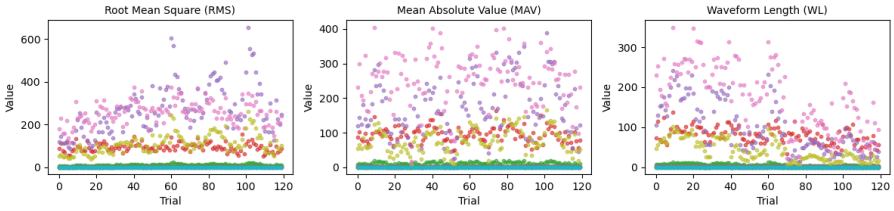


Figure 1.1
RMS, MAV and WL features across 120 trials, colored by channel.

1.2 Splitting data, perform classification and evaluate

Data were split at the repetition level. For each movement, 7 repetitions were randomly selected for training, 2 for validation, and 1 for testing to ensure class balance within each set. The training set is used to train the model and optimize hyperparameters, the validation set to assess performance on unseen data and compare different feature selection methods, and the test set to evaluate the best model performance.

We initially used all features and performed hyperparameter tuning over $C = [0.1, 1, 10, 100]$, kernel $= [\prime\text{linear}\prime, \prime\text{ rbf}\prime, \prime\text{ poly}\prime]$, and degree $= [3, 4, 5]$ (for the polynomial kernel) using 5-fold cross-validation, selecting the combination that maximized the accuracy score. The best combination was rbf with $C = 1$, confirmed on the validation set. The final model achieved an accuracy of 0.96 (and a F1 of 0.96 as well) on both validation and and an accuracy of 0.83 (and a F1 of 0.78) on the test set, indicating strong performance.

The F1 score balances precision (the proportion of correctly predicted positive samples) and recall (the proportion of actual positive samples correctly identified) and is particularly useful when class distributions are

imbalanced, as accuracy can be misleading if some classes dominate. However, since we ensured that all splits contained balanced classes, accuracy alone provides a reliable performance measure in our case.

1.3 Feature selection and PCA

We used three feature selection methods: Recursive Feature Elimination (RFE) with a linear SVM (step 0.1), Principal Component Analysis (PCA), and the top 'n' features by mutual information with the target. After feature selection, we re-tuned hyperparameters with 5-fold cross-validation on the training set. For each method, we select the top $[5, 10, 15, 20, 25, 30]$ features/components and report here only the final selected features that did not significantly reduce performance.

Table 1.1 summarizes the performance of those models across different feature selection methods. For most methods, performance does not drop significantly, reflecting that many features are redundant or weakly informative. Given our relatively small validation set and the high original model performance, it is possible that some feature selection techniques could even slightly improve performance. Specifically, PCA achieves near-original performance with only 10 components, and RFE performs similarly with the top 15 features. In contrast, selection based on mutual information requires a larger number of features to approach the performance of the original model.

Table 1.1
Performance of different feature selection methods and models.

Features	Best Model	Val accuracy	Test accuracy
All	RBF / C=1	0.96	0.84
RFE top 15	RBF / C=1	0.88	1
PCA 10 components	RBF / C=10	0.92	0.92
Mutual Info top 25	Linear / C=10	0.88	0.75

2 Part Two: Generalization across subjects

2.1 Pre-processing and Feature Extraction

The EMG signals from all 27 subjects were visually inspected and found to be already rectified and low-pass filtered below 50 Hz. Therefore, no additional filtering was applied. Pre-processing consisted of segmenting the data into trial-specific windows based on stimulus and repetition labels. For each subject, the data were divided into 12 stimulus classes with 10 repetitions per class, and these windows were used directly for feature extraction. From each EMG window, a set of time- and frequency-domain features was extracted per channel, including RMS, MAV, WL, SSC, IEMG, VAR, WAMP, HISTLOG, MDWT, and mean PSD.

To examine inter-subject variability, feature distributions were compared across the 27 subjects. Figure 2.1 shows selected features for two example channels. While overall feature patterns are consistent across subjects, substantial inter-subject variability is observed, particularly for IEMG and WAMP, which span markedly different value ranges as reflected by the y-axis scales. In contrast, PSD and VAR show comparatively lower variability across subjects.

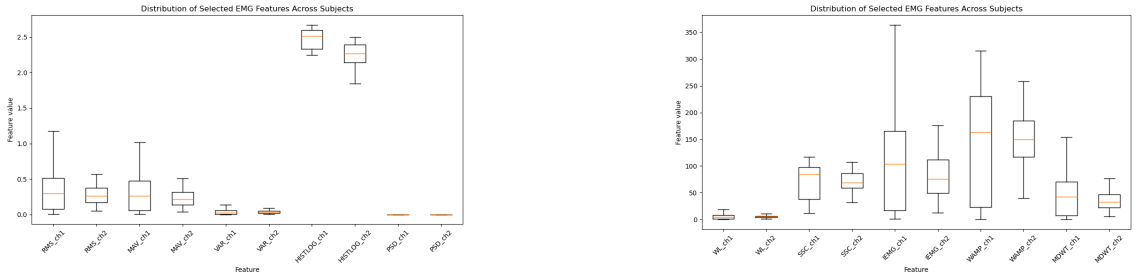


FIGURE 2.1
Distribution of selected EMG features across subjects for two example channels.

2.2 CROSS-SUBJECT CLASSIFICATION AND LOSO EVALUATION

To evaluate cross-subject generalization, the same RBF-kernel SVM classifier as in Part 1 was trained on data aggregated from 26 subjects and tested on a held-out subject. This procedure was repeated using leave-one-subject-out(LOSO) cross-validation, such that each of the 27 subjects was used as the test set once. Classification performance was evaluated using accuracy, as all stimulus classes were balanced and equally represented in the dataset.

Across all LOSO folds, the mean cross-subject accuracy was 0.44 ± 0.13, with large variability across subjects. In contrast, within-subject classification using the same repetition-based splitting strategy as in Part 1 achieved consistently higher performance. As shown in Fig. 2.2, the pronounced performance gap between within-subject and cross-subject decoding indicates limited generalization to unseen subjects. While the classifier achieves robust performance when trained and tested on data from the same subject, its accuracy decreases substantially under LOSO evaluation, reflecting differences in muscle activation patterns, signal amplitudes, and physiological characteristics across individuals.

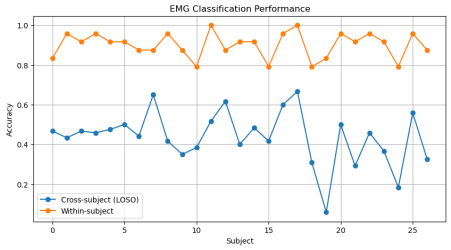


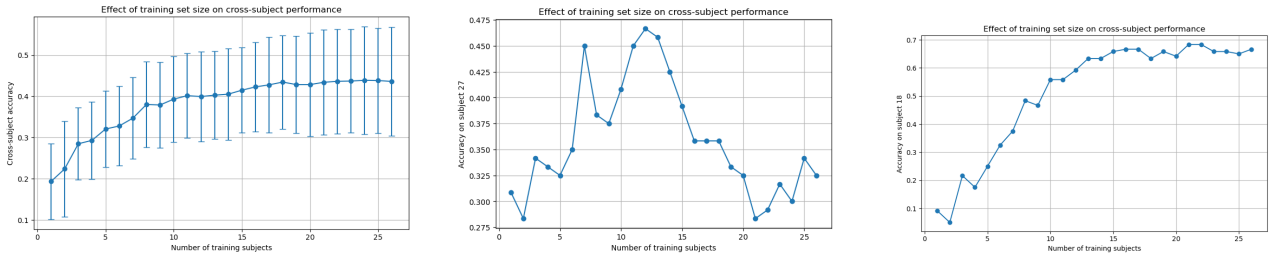
FIGURE 2.2
Comparison of classification accuracy for within-subject and cross-subject evaluation across the 27 subjects.

2.3 EFFECT OF TRAINING SET SIZE ON CROSS-SUBJECT PERFORMANCE

To study the impact of training set size on cross-subject generalization, the classifier was trained using an increasing number of subjects and evaluated on a held-out test subject. This procedure was repeated for all subjects, and the mean accuracy and standard deviation across test subjects were computed for each training set size. The resulting performance is shown in Fig. 2.3a.

Cross-subject accuracy improves substantially as more subjects are included in the training set, particularly when increasing from a small number of subjects. Beyond approximately 15 training subjects, accuracy improvements become marginal, while variability across subjects shows no further systematic reduction and exhibits slight fluctuations.

From a practical perspective, training with too few subjects leads to poor and unstable generalization, whereas increasing the training set size improves performance at the cost of higher computational and data requirements. Therefore, when designing a subject-independent decoder, it is important to balance accuracy, generalization, and computational load when selecting the size of the training dataset.



(A) Mean ± STD across subjects **(B)** Test subject 27 **(C)** Test subject 18

FIGURE 2.3
(a) Mean accuracy and standard deviation across all subjects. (b–c) Example accuracy trends for two individual subjects.

3 Part Three: Regression for joint angles

3.1 PRE-PROCESSING AND SLIDING WINDOWS

The dataset consists of 16 EMG channels with 2,292,525 samples each, corresponding to 1146.3 seconds of recording.

We inspected the power spectral density (PSD) of the signals and confirmed that power-line noise at 50 Hz and its harmonics had already been removed (Figure 3.1). The PSD also showed that most signal energy lies within the 5–500 Hz range. To ensure consistency and suppress any residual out-of-band noise, a 5–500 Hz Butterworth band-pass filter was applied to all channels. The EMG signals were rectified to ensure all values were positive.

The dataset was divided into training, validation, and test sets using a 70-20-10 split based on repetitions, in order to preserve temporal dependencies and ensure meaningful sliding windows. 70% train (repetitions 1, 3, 5, 6, 8, 9, 10), 20% validation (repetitions 2, 7), and 10% test (repetition 5).

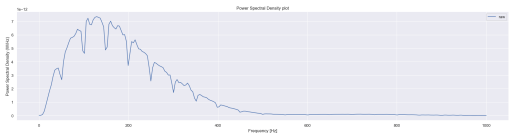


FIGURE 3.1
Power spectral density of a representative EMG channel.

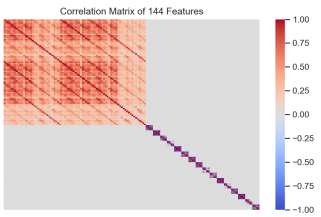


FIGURE 3.2
Correlation matrix across 144 features.

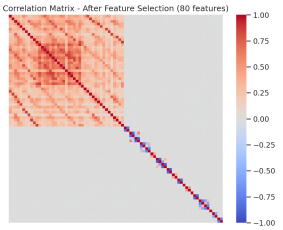


FIGURE 3.3
Correlation matrix across 80 features.

The envelopes of each data split were computed using a moving average. The dataset was then segmented into overlapping sliding windows, with a window length of 128 ms and an increment of 50 ms, resulting in 60% overlap between consecutive windows (EMG window length: 256 timepoints, Incremental window length: 100 timepoints). The choice of window length follows the original study (Krasoulis et al. 2019), and ensures that each window contains 142 time points. This length captures sufficient temporal information while maintaining substantial overlap between windows, which is beneficial for modeling temporal dependencies in EMG signals.

3.2 FEATURE EXTRACTION

Based on the lab exercises and the original study (Krasoulis et al. 2019), we extracted 11 features per channel: Mean, Standard Deviation, Maximum Amplitude, Wilson Amplitude, Waveform Length, Log-Variance, Slope Sign Change, and 4th-order Auto-Regressive (AR) coefficients. With 16 channels, this results in 176 feature columns; after removing constant columns, 144 remain. Features were standardized using z-score normalization.

The correlation matrix shows positive correlations among certain features (e.g., {`logvar_2, std_2`}, {`max_11, mean_11`}, {`ar1_1, ar2_1`}) corresponding to the same EMG channel, as expected, since these features capture related aspects of the same underlying signal dynamics (Figure 3.2).

To reduce multicollinearity, we removed one feature from any pair with correlation > 0.90 or < -0.90, keeping the threshold high to preserve important information. For example, AR coefficients are not redundant even if they share high correlation because each coefficient captures a different lag in the signal and contributes uniquely to the temporal structure. After this selection, 80 features remained (Figure 3.3).We observe diagonal lines in the correlation matrix, indicating that different features for different channels are highly correlated, which matches our expectations.

3.3 Regression on Kinematics

We fitted the SVR model on the training set using two kernels ("linear" and "rbf") and five regularization parameters ($C = 0.01, 0.1, 1, 10, 100$). Table 3.4 summarizes the resulting mean squared errors (MSE) for each configuration.

Kernel / C	0.01	0.1	1	10	100
linear	694.53	586.32	579.65	578.98	584.07
rbf	1042.09	932.74	550.88	352.97	303.82

Figure 3.4
SVR performance (MSE) comparison

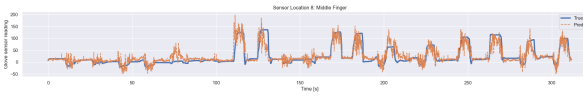


Figure 3.5
Middle finger results visualization

The optimal SVR parameters were: $C = 100$, kernel = rbf. Validation: MSE = 303.82, R^2 = 0.68. We also tried to use PCA (keeping 35 components) on the features and check the performance with the same model ($C = 100$, kernel = rbf). The results (MSE = 402.30, R^2 = 0.57) indicated that the performance is significantly impacted. We hence decided to give up on PCA.

Figure 3.5 shows the true versus predicted angles of the middle finger on the validation set. In this example, the model accurately captures the middle finger joint angle, although prediction noise is present. The model produces comparable predictions for the other fingers, with values within the same range.

3.4 Evaluation and Discussion

MSE penalizes large errors and is optimal for Gaussian noise assumptions, but R^2 is more appropriate here as it normalizes by label variance. We use $R^2 = 1 - \text{MSE}/\text{Var}(y)$ to assess test performance. For the test set, we obtained MSE = 325.76. With R^2 = 0.70, the model captures most of the variance in the kinematic signals, indicating strong predictive performance given the complexity and noise in EMG data.

Table 3.6 shows R^2 scores for each finger, revealing slight variation in model performance. The middle and little fingers achieved high scores, while the index and ring fingers had the lowest, and the thumb finger was intermediate. This aligns with prior findings of the original study (Krasoulis et al. 2019): decoding accuracy depends on signal clarity and selective muscle activation. The middle finger is largely independent with distinct tendon compartments, producing strong EMG patterns. The little finger is moderately independent, while the ring finger is, for example, biomechanically coupled to adjacent fingers, leading to mixed EMG signals. Thumb motion relies on intrinsic hand muscles not captured by forearm sEMG, and extrinsic muscles overlap with wrist muscles, causing crosstalk, therefore has lower results than the middle finger. These results complement previous reports of higher decoding accuracy for little/middle fingers and lower accuracy for other fingers.

Figure 3.6
R^2 values for each finger sensor location.

Sensor / Finger	Thumb	Index	Middle	Ring	Little
R^2	0.700	0.673	0.743	0.675	0.698

4 Appendix

4.0.1 TEAM CONTRIBUTIONS

Name	Tasks / Contributions
Angana Mondal	Part 3
Arnault Stähli	Part 3
Heliya Shakeri	Part 2
Khushi Singh	Part 1
Pamela van den Enden	Part 1-2

TABLE 4.1
Summary of contributions.

Bibliography

Atzori, Manuella, Arjan Gijsberts, Claudio Castellini et al. (2014). 'Electromyography data for non-invasive naturally-controlled robotic hand prostheses'. In: *Scientific Data* 1.1, p. 140053. DOI: 10.1038/sdata.2014.53.

Atzori, Manuella, Arjan Gijsberts, Ilya Kuzborskij et al. (2015). 'Characterization of a benchmark database for myoelectric movement classification'. In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 23.1, pp. 73–83. DOI: 10.1109/TNSRE.2014.2328495.

Krasoulis, Alexandros et al. (2019). 'Effect of user adaptation on prosthetic finger control with an intuitive myoelectric decoder'. In: *Frontiers in Neuroscience* 13. DOI: 10.3389/fnins.2019.00891.