Machine Learning classification and Regression using EMG

Group Members: Tristán Fonquerne Torres, Nicolás Reátegui Rodríguez,
Emanuele Rimoldi, Imane Jennane

Introduction

Electromyography (EMG) signals are essential for decoding human motor intentions, with applications in assistive technologies such as robotic prostheses and rehabilitation systems. These signals, generated by skeletal muscle activity, can be analyzed to recognize movements or predict joint angles, enabling precise control of assistive devices. The NinaPro Dataset [1], a widely used resource for EMG-based gesture recognition, forms the basis of this project. Subject 2's data from Dataset 1, Exercise 1 (`S2_A1_E1`) is used for Parts 1 and 2, focusing on single-subject classification and cross-subject generalization, respectively. In Part 3, data from Subject 1 of Dataset 8 (`S1_E1_A1/A2/A3`) is used to implement regression models for predicting joint angles, aiming to enhance the control of robotic hands and achieve finer movements, such as half-closed positions.

1 Single subject classification

1.1 Data Visualization and Preprocessing

We use the NinaPro Dataset 1, specifically Subject 2 and Exercise Set 1 (`S2_A1_E1`). It comprises EMG signals recorded across 10 channels, totaling 100,686 samples. The dataset is complete, with no missing samples or corrupted trials. A detailed inspection of variance and mean across all channels confirmed their validity, as there are no constant channels. This ensures the integrity of the dataset, allowing us to include it in its entirety for analysis. We stratify samples by stimuli and repetitions, then apply a moving average filter for smoothing. This step effectively reduces noise and improves the quality of our features. We determine the optimal window length by assessing different window lengths' smoothing and noise reduction. We observe that a window of around 25 measurements effectively reduces noise without compromising the shape of the EMG signal.

1.2 Splitting the Dataset

We split our dataset to train and test machine learning classifiers. The training and validation sets were combined to maximize the data available for model training, optimizing model performance by allowing it to better capture underlying patterns and improve generalization. The test set remained separate and was reserved for final evaluation, ensuring unbiased performance metrics and maintaining the integrity of the evaluation process. We used stratified splitting to ensure a balanced representation of all 12 movement classes across subsets, which is critical to prevent overfitting and ensure the model generalizes to unseen data from different stimuli we aim to recognize. Each stimulus-repetition combination was treated as a distinct unit, maintaining independence, preventing fragmentation between subsets and avoiding test set contamination. For hyperparameter optimization, we used Grid Search with **Stratified K-Fold Cross-Validation** on the combined training and validation sets. Stratified K-Fold ensures proportional label representation across folds while redistributing the data during each iteration. This approach optimizes the Random Forest classifier using a more comprehensive dataset and avoids under-utilizing available data while maintaining the integrity of the test set.

1.3 Feature Extraction and Analysis

Six EMG features are extracted: Mean Absolute Value (MAV), Standard Deviation (STD), Maximum Absolute Value (MAXAV), Waveform Length (WL), Root Mean Square (RMS), and Slope Sign Changes (SSC), capturing signal energy, complexity, and transitions. As shown in Figure 1, MAV and RMS exhibit consistent values across channels, with peaks observed in Channel 9, reaching up to approximately 2.0. This consistency highlights their role in representing signal energy. STD and MAXAV follow a similar trend but display slightly larger peaks, with STD reaching values of 1.2 and MAXAV peaking at 4.5, particularly in Channel 9. Waveform Length (WL) shows a broader range of values between 5 and 35, with noticeable variations in Channel 9. In contrast, SSC presents a distinct pattern compared to the other features, with values ranging widely from 0 to 150 and varying significantly across channels.

The observed discrepancies in SSC values suggest higher sensitivity to signal transitions, possibly caused by noise or abrupt changes in muscle activity. Additionally, the range of values across the first four features (MAV, STD, MAXAV, and WL) remains relatively limited compared to the broader ranges seen in WL and especially SSC. The consistency observed across trials for most features confirms feature stability, while the channel-specific discrepancies can be attributed to anatomical variability, electrode placement, and differences in muscle activation.

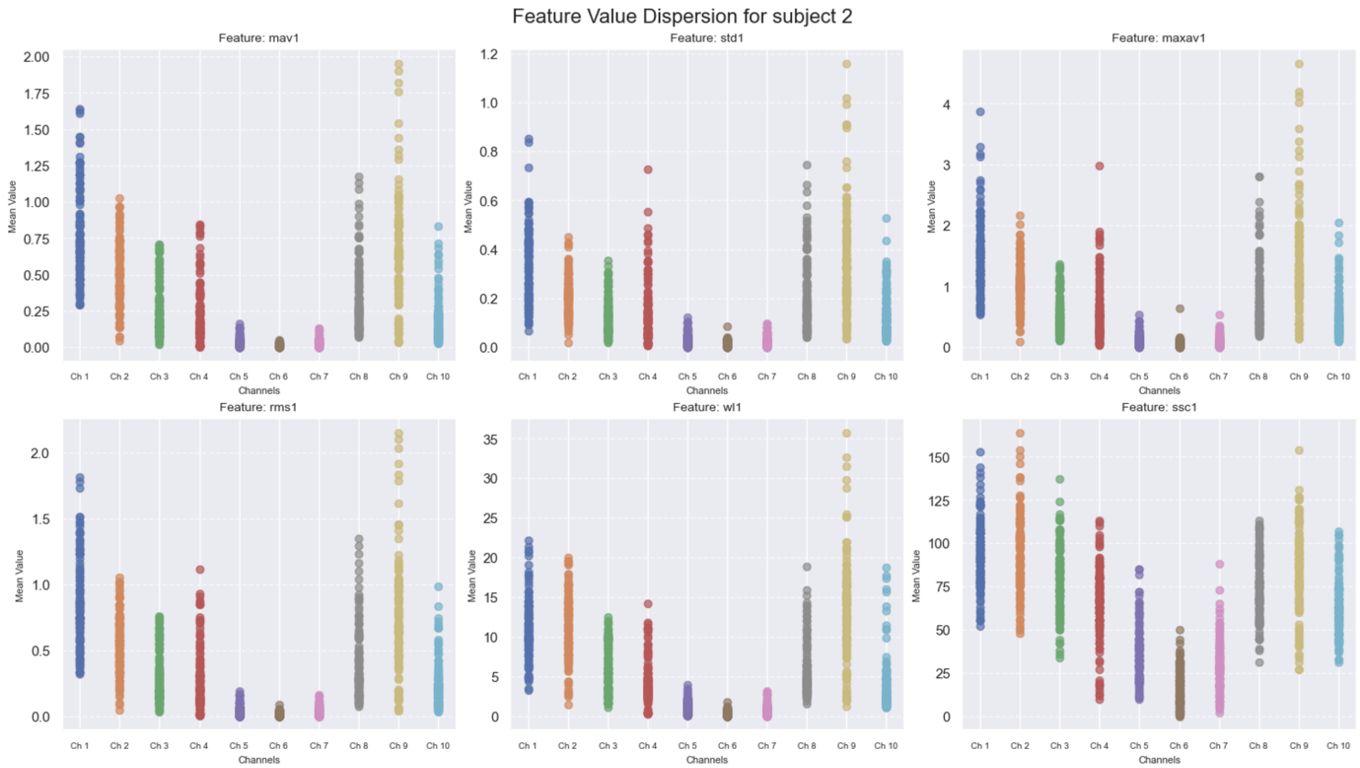

Figure 1: Feature value dispersion across EMG channels for the six extracted features.

1.4 Classification

We train a Random Forest Classifier to predict movement classes and obtain a **test accuracy of 87.50%.** After hyperparameter optimization, including grid search over the number of estimators ($n_{\text{estimators}} = 300$), maximum tree depth (max_depth $= 10$), and feature selection methods, the test accuracy improved to **91.67%.** The best-performing class is **Class 1**, achieving perfect precision, recall, and F1-score. In contrast, Class **6** shows the lowest recall, indicating that some samples in this class are misclassified. The confusion matrix highlights a small number of misclassifications, primarily between similar movements, while the ROC-AUC score of **0.9962** confirms the model's strong discriminative power. Given its interpretability and relevance in balanced multi-class classification tasks, accuracy is used as the primary evaluation metric. The improvements in accuracy following hyperparameter tuning validate the effectiveness of the selected features and classifier configuration. Overall, the performance metrics demonstrate a balanced and robust classification across all 12 movement classes, with hyperparameter tuning significantly enhancing the model's accuracy and reliability.

1.5 Feature Selection and Dimension Reduction

We use both Mutual Information and SHAP (SHapley Additive exPlanations) for feature selection. Mutual Information identifies features with the highest statistical dependence on the target variable, while SHAP provides a model-agnostic explanation of feature importance. Both methods consistently identified MAV and RMS as top contributors. Furthermore, utilizing only the top-ranked features (approximately 20 for both methods employed) reduces the dimensionality while retaining predictive power. Additionally, the optimized models achieve comparable or slightly improved accuracy (95.83%) with respect to training on the full feature set. Therefore, feature selection can enhance the model's interpretability without compromising performance.

2 Generalization across subjects

In this section, we assess whether classifiers generalize across various subjects. To that end, we use all subjects from the Ninapro 1 dataset, for a total of 27 subjects.

2.2 Feature Extraction and Analysis

We preprocess the data and extract features (MAV, std, MAXAV, RMS, WL, and SSC) using the same pipeline as in Part 1. We then extract the features and store them separately for each subject. After that we create our train and test dataset by concatenating subjects in different ways to verify the possibility of generalization across different subjects for classifications. When observing the features, we can see that all features but the SSC have the same order of magnitude. Additionally, there are similarities between several repetitions of the same movement as well as some discrepancies w.r.t. repetitions and subjects.

2.3 Extraction of one subject and comparison

For this analysis, we extract one subject (subject 10) and train two random forest classifier models. The first model is trained with data from all subjects except subject 10, and the second includes a subset of samples from the reserved subject. We compare their performances using the accuracy and confusion matrix. We train a Random Forest Classifier and obtain the confusion matrix shown in Fig. 3. Our model yields 45% accuracy for the inter-subject classification and 87.5% when training another model with a train/test split version of the dataset from the subject 10.

When training with other subjects and testing with subject 10, some movements have good identification (stimuli 1, 2 and 4) while the others have variable recognition (stimuli 5, 6 and 10 always get wrongly classified). Performances for inner-subject classification are way better than for inter-subject as shown in Fig. 3, this can have many explanations, we will perform cross-validation before concluding.

2.4 Cross-validation and Rotating subjects

We now use the same methods and rotate the test subject. For each subject, we compute the confusion matrix and the accuracy. Similar to the previous observations, we obtain better results for intra-subject than inter-subject classification. We provide a summary of the accuracy for both cases in Table 1.

The low accuracy for inter-subject classification shows the difficulty of generalizing across subjects Although we train on a large number of subjects, we believe increasing this number can solve this issue. There are many factors contributing to the data differences between subjects including biological differences and EMG sensors. The latter is particularly important as the position of the sensors determines the signals recorded and the channels required for a proper comparison. Given that sensors are manually

Metrics	Inter-Subject	Intra-Subject
Mean	33.8	91.2
Standard Deviation	10.87	4.99
Maximum	60	100
Minimum	15.83	83.33

Table 1: Accuracy comparison for intra and inter-subject classification (values in %)

placed, there is high variability across subjects. Additionally, to produce comparable data, we tried standardizing the features by channels across all subjects; the same standardization values are used for the test set. Despite our efforts, this method does not show significant improvements.

2.5 Varying number of training subjects

This analysis aims to assess the impact of increasing the number of subjects on accuracy. We sequentially include all subjects in training except subject 10 which was reserved for testing (Figure 2). Increasing the number of subjects leads to significant improvements from 15% to 48% accuracy. Nevertheless, the maximal value is still low (50%) and the accuracy seems to reach a plateau after adding around 15 subjects. We can see that there is some kind of generalization between subjects but this has some limitations as we have seen in the previous tests. We can note that subject 10 got particularly good results but other subjects had less significant improvement when adding the other subjects. This difference in classification might come from some similarities between some subjects or the signals that we have obtained, in the graph we can see a big spike after adding subject 14, this subject might have good classification results with subject 10.

Figure 2: Accuracy of Random Forest models trained on an increasing number of subjects

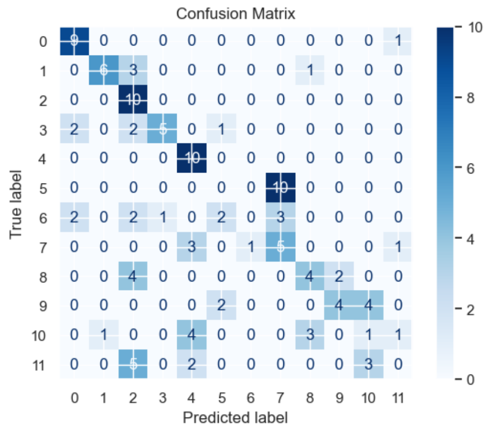

Figure 3: Confusion matrix for inter-subject testing

3 Regression for joint angles

In this section, we implement a regression model to predict joint angles using the NinaPro Dataset 8. We train, test, and validate our models on subject 01 using datasets A1, A2, and A3 respectively.

3.1 Preprocessing

We begin by importing the datasets and extracting the channels for the required glove angles. Initial visualization shows an un-rectified signal with only powerline noise filtering. The latter is visible through power decreases associated with the main frequency (50Hz) and its harmonics. We apply a `bandpass` filter between 5 and 500 Hz and rectify the EMG data for all datasets (dataset 1, dataset 2, and dataset 3). Furthermore, we use a `bandstop` filter to remove spikes appearing on specific channels and frequencies (147-149). Then, we apply a 1D convolution over 400 time points to produce the EMG envelope. As we use different datasets to train, test, and validate our models, we have no need for specific splitting; this also avoids the inherent issues when dealing with time series data.

3.2 Feature Extraction

We extract features over sliding windows. Given the number of time points selected and the provided reference [1], we use a 128 ms window. Furthermore, we have a step of 50ms to have 60% of overlapping data. After that, we perform the train/test split using datasets 1 and 2, avoiding overlap issues.

3.3 Extracting features and correlation

After this, we extract the features and plot the pearson correlation (Figure 6). We observe that the MAX feature has a maximum correlation with the mean value its associated channel. We believe that the presence of outliers or spikes severely influences the resulting mea leading to a high correlation. Furthermore, there is an overall high correlation between mean and max values across channels.

3.4 Performance visualization and analysis

We visualize the performance of our regressor by comparing the predicted and true values for each joint angle (Figure 5). We observe a linear relationship between both values indicating a good fit. Furthermore, we use the RMSE to provide a quantitative evaluation of our model. This metric is useful for regression tasks as it allows us to compare the error using the same units as the input. We believe our model is moderately good as we obtain a mean RMSE of 5.1.

3.6 Single finger analysis

Finally, we evaluate the RMSE for each finger and observe a significantly larger value for the index finger. As this finger has the largest range of values from -20 to +60, this leads to a large RMSE. One of the main drawbacks of the RMSE is that comparison across different scales is not straightforward.

4 Conclusion

In conclusion, we have demonstrated that it is possible to classify and recognize different movements using a small set of features. Furthermore, the classifier achieves high-performance scores. We also explored Mutual Information and SHAP methods for feature selection. We then attempted to generalize this process to inter-subject classification. Despite these efforts, we observed that the model performed significantly better when trained and tested on data from a single subject. Finally, we developed a regressor to predict joint angles of the fingers. Our results show that it is possible to accurately predict finger joint angles from EMG signals with an overall acceptable precision level.

References

[1] S. Vijayakumar & K. Nazarpour A. Krasoulis. Effect of user practice on prosthetic finger control with an intuitive myoelectric decoder. *Frontiers Neuroscience*, 13, 2019.

Appendix A: Additional figures

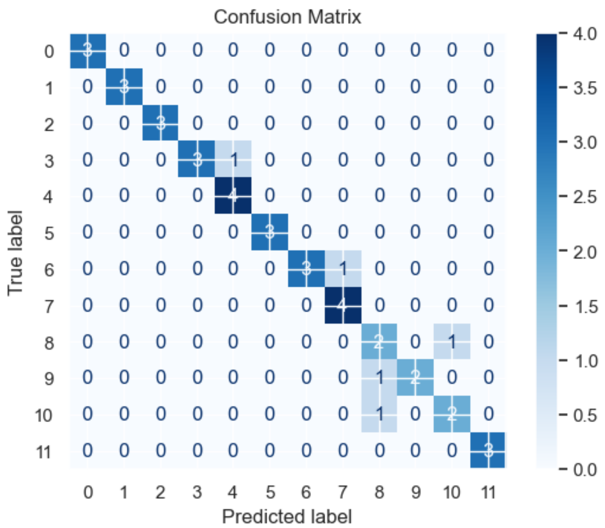

Figure 4: Confusion matrix for intra-subject testing

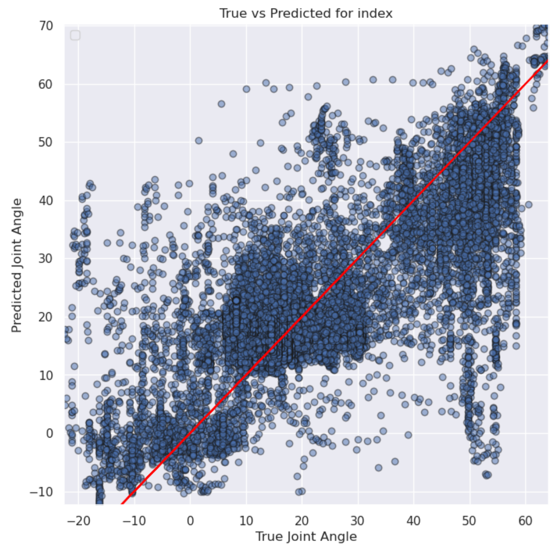

Figure 5: True vs predicted joint angles for the index

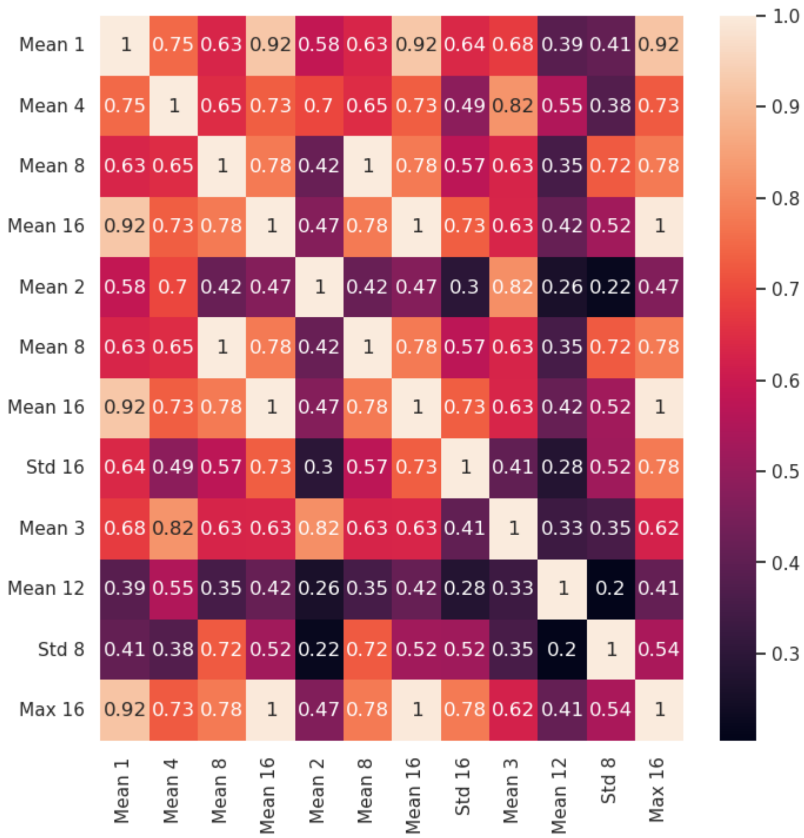

Figure 6: Feature correlation for the joint angle regression