Executive Summary: Discrete Gesture Recognition

Brian R. Mullen, Carrie Clark, Revati Jadhav, Philip Nelson, Sero Toriano Parel

Github Repository

Introduction

Smart wristbands have the potential to let people control devices with subtle hand gestures just by “reading” the wearer’s muscle signals while being non-invasive. However, everyone’s muscle signals are a bit unique. Since surface electromyography (sEMG) signatures show significant inter-user variability, our main challenge is personalization: ensuring the model generalizes reliably to unseen actions by the same individual.

We are using the discrete gesture sEMG data from the generic neuromotor interface dataset, encompassing 135,299 samples from 100 participants (Kaifosh et al. 2025 Nature) who performed 9 gestures. While Kaifosh et al. used deep learning for “out of the box” gesture recognition, we sought out a simpler approach, asking: Can engineered features enable a personalized, interpretable model for gesture recognition?

Stakeholders and KPIS

Stakeholders are not only the end users and consumers of the device who want to be able to control devices with subtle hand gestures, but also commercial interface developers such as Meta Reality Labs who require robust and scalable models to deploy in hardware.

Our primary metric is the F1 macro score, which we have chosen because our data is imbalanced across gesture categories. Secondary metrics include accuracy and classification error rate (CLER).

Preparing the Data

Data Source: Generic neuromotor interface dataset (Kaifosh et al. 2025 Nature) with 135,299 samples, 160 features, 100 users, 9 gestures. Limitations include inter-user variability and systematic confusion patterns between similar gestures.

Signal Alignment: The data was annotated with when each user was prompted to perform the task, but did not indicate when the gesture was enacted. As such, we first had to align the muscle activity between each trial. The approach we took was to find a large event that achieved 3 standard deviations above baseline that was close to the prompted gesture, assuming that the large event was a result of the prompted gesture.

Feature Extraction: We extracted 10 types of features, leading to 160 total features across the 16 channels. These metrics are standard ways of analyzing EMG signals, including root mean square (RMS), mean absolute value (MAV), characteristics of the fast Fourier transform (FFT), and threshold crossings. From these features, we removed outliers using Gamma distribution detection (2.9% of data) and handled missing values via median imputation.

Methods and Modeling

Following the personalization approach, each user's data was split into 80% training data and 20% testing data, stratified by both gesture and trial stage (note: in each stage the user held their arm in a specified posture, for example thumb_swipes_static_arm_raised or pinch_release_dynamic_vertical_arm_translation). Models were then trained on each user separately.

Feature Engineering: To rank the importance of the 160 features, we used a random forest model to measure feature importance. We observed that RMS features were the most predictive and that channels 4 and 5 were highly discriminative. To address collinearity, we used a threshold correlation of 0.9 to remove redundant features from the top 80 features ranked by random forest analysis. This resulted in a more compact set of 37 selected features, yielding a 77% reduction in features.

Modeling: To address class imbalance and feature dimensionality, we evaluated both the selected 37-feature set versus the full 160-feature set across multiple model families using 5-fold stratified cross-validation within each user. Models tested included logistic regression with L2 regularization with and without balanced class weighting, XGBoost, and random forests. We used a dummy classifier as a simple baseline. In our cross validations, the best personalization performance was attained by logistic regression with L2 regularization on the 37-feature subset, scoring a mean F1 Macro of 0.7164 and accuracy of 0.7340.

Results

Holdout testing revealed a generalization gap: mean F1 Macro dropped to 0.3977 and accuracy to 0.4678. Confusion between release gestures jumped to 28%. While the model still beat random chance, performance varied widely, with 65% of users scoring below 0.5 accuracy and only 4% exceeding 0.7. To investigate this drastic drop in performance, we found that users with more data achieved up to 15% higher accuracy.

Future Directions

Improvement could be attained on time window selection. Our selection of 100 ms for each gesture worked well for sustained gestures (i.e., press and release), but more transient gestures such as swipe and click gestures may only require only 10-50 ms.

Future work could explore adaptive models such as LSTMs, deep learning, or template matching to address the heterogeneity in performance and to improve generalization to novel gestures within individuals.