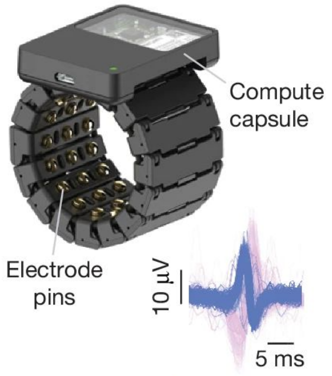
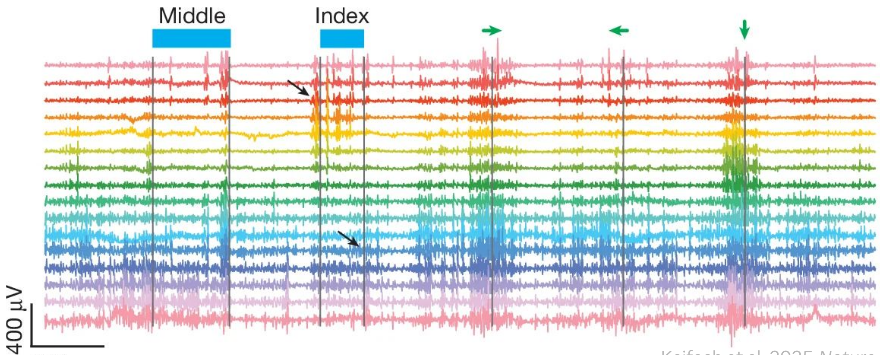
Personalized Gesture Recognition

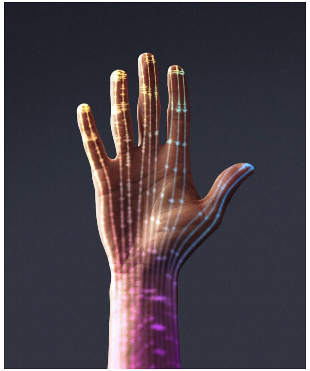
Brian R. Mullen, Carrie Clark, Revati Jadhav, Philip Nelson, Sero Toriano Parel

Erdős Institute Data Science Boot Camp Fall 2025 Project





Kaifosh et al. 2025 *Nature*



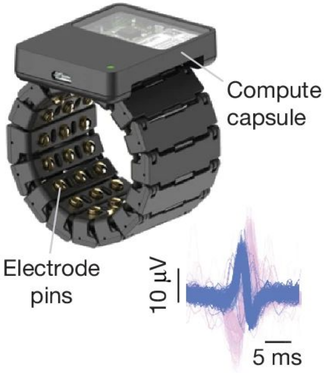
Introduction

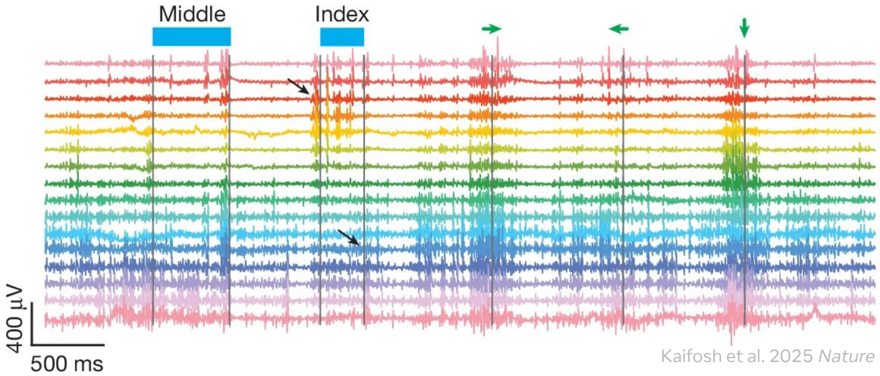
Article | Open access | Published: 23 July 2025

A generic non-invasive neuromotor interface for human-computer interaction

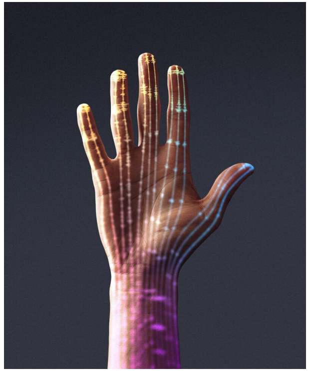
Patrick Kaifosh, Thomas R. Reardon & CTRL-labs at Reality Labs

Nature **645**, 702–711 (2025) | Cite this article

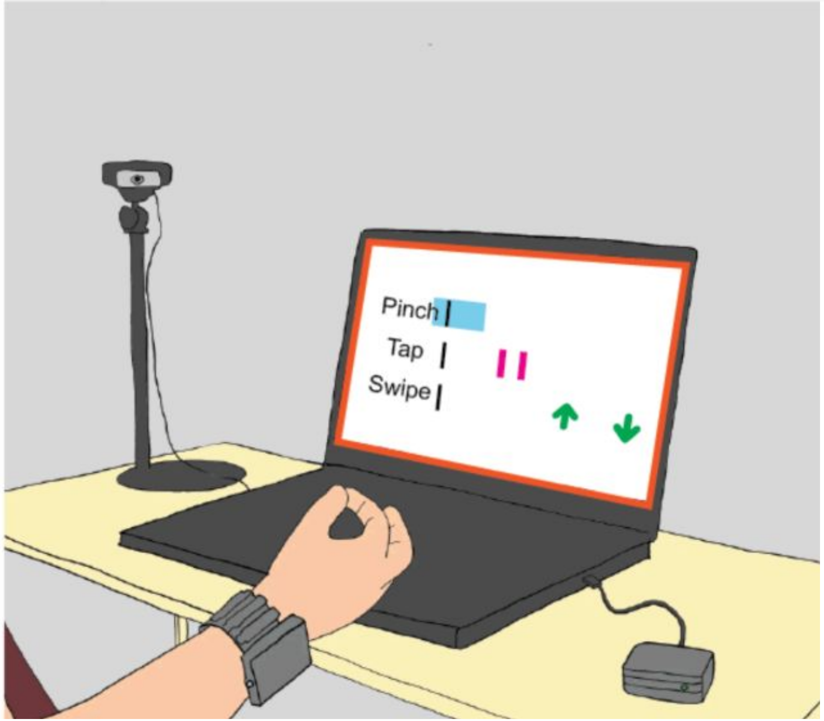


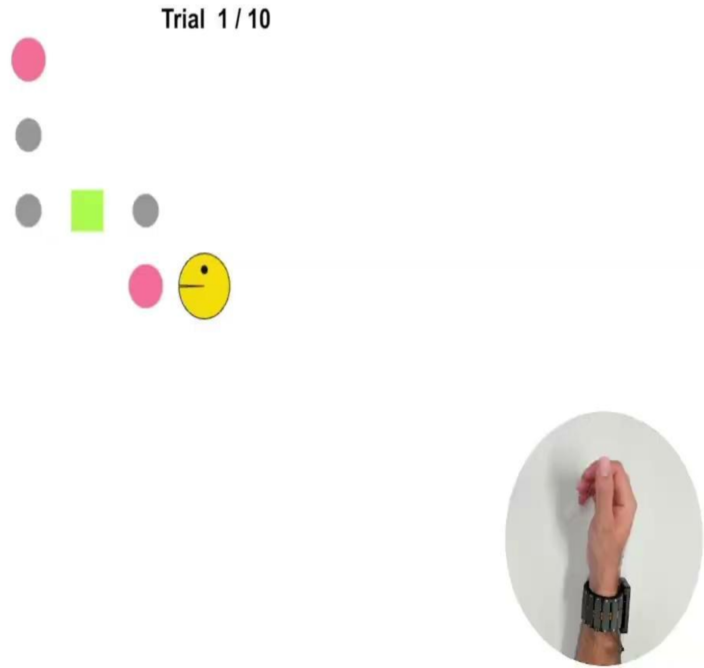


Kaifosh et al. 2025 *Nature*



Setup of wristband and gestures





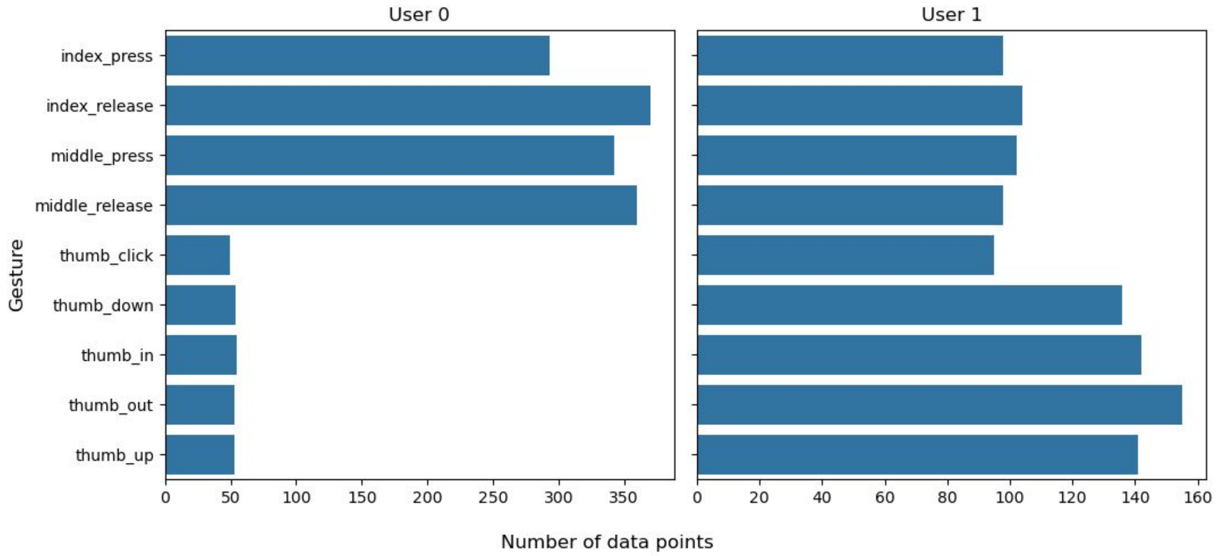
Project problem

Can we build a robust, personalized classifier to accurately distinguish between 9 discrete hand gestures using non-invasive sEMG signals, overcoming high inter-user signal variability?

Data Structure

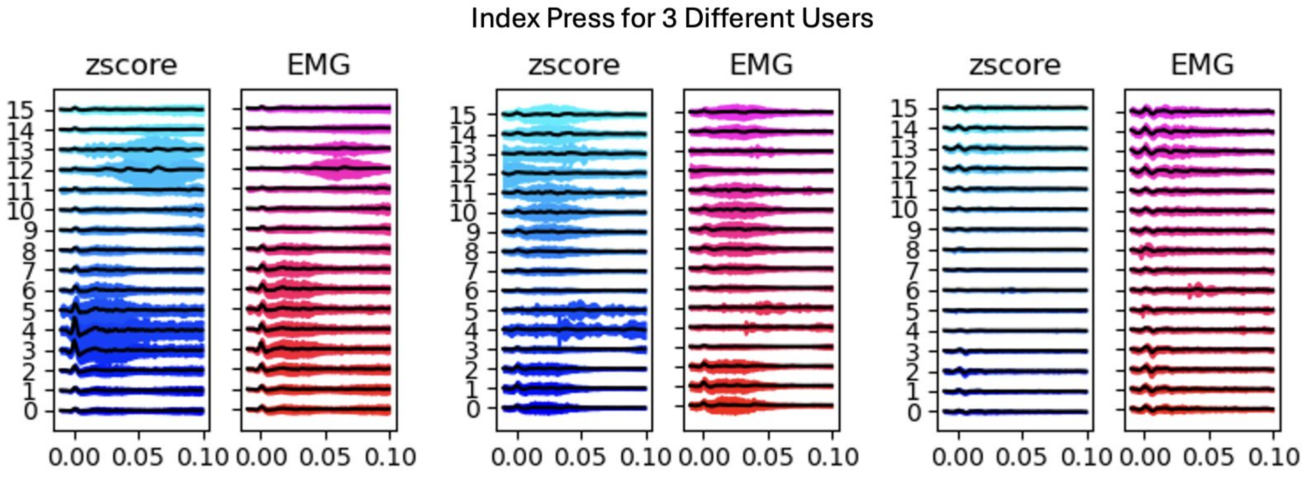
Source: discrete gestures dataset, generic-neuromotor-interface from Reality Labs

- 100 participants
- 9 gestures, performed up to 400 times during various stages (e.g. thumb_swipes_static_arm_raised, pinch_release_dynamic_vertical_arm_translation, …)



Personalization Approach

Challenge: emg data varies significantly between users

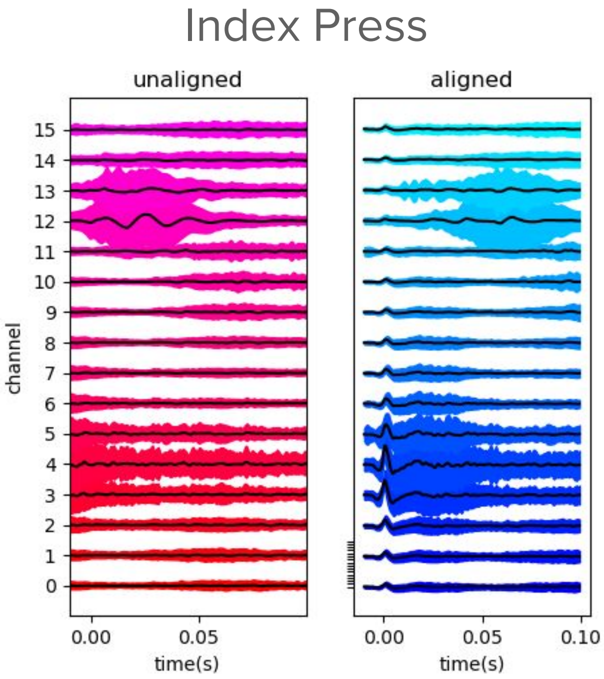


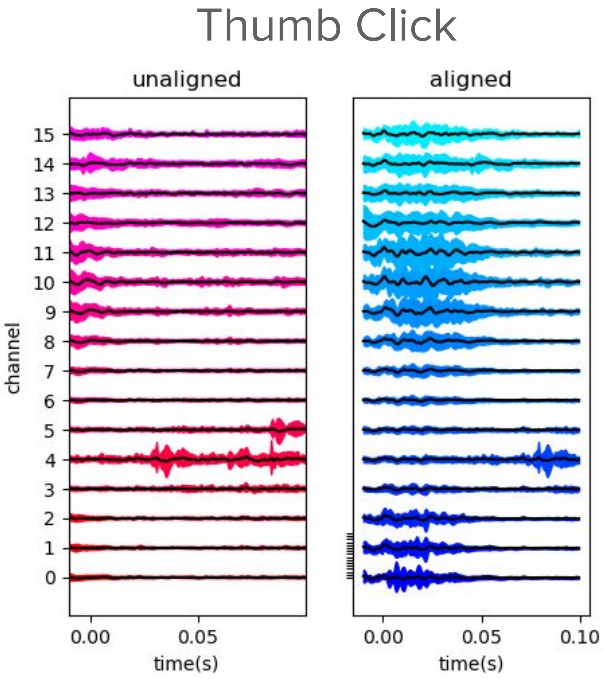
Personalization Approach:

- Each user's data is split into an 80\20 train test split, stratified by stage and gesture
- Model is trained for each user separately

Alignment of signals

Before extracting features, we aligned the emg data by aligning to the closest large event to the prompted gesture





Feature Engineering

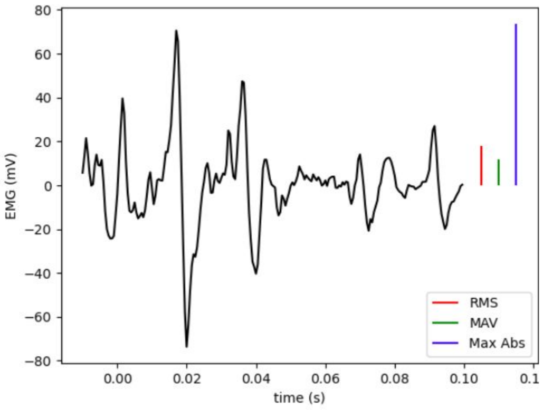
Features extracted from signal data from each of the 16 channels:

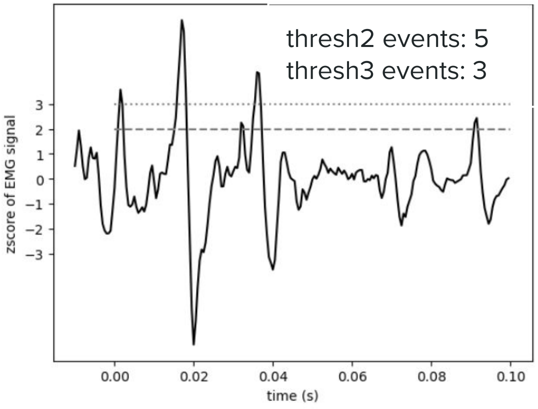
- Root mean square
- Mean Absolute Value
- Maximum absolute value

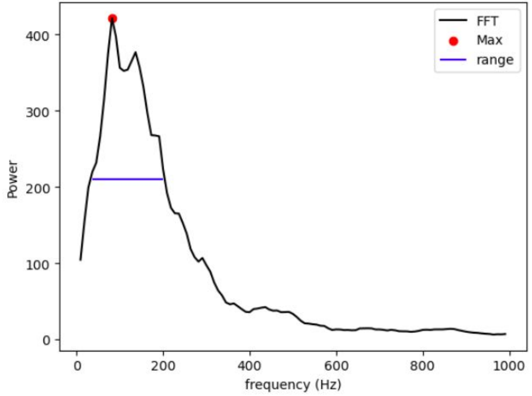
- thresh2-events
- thresh3-event

- FFT maxpower
- FFT peak freq

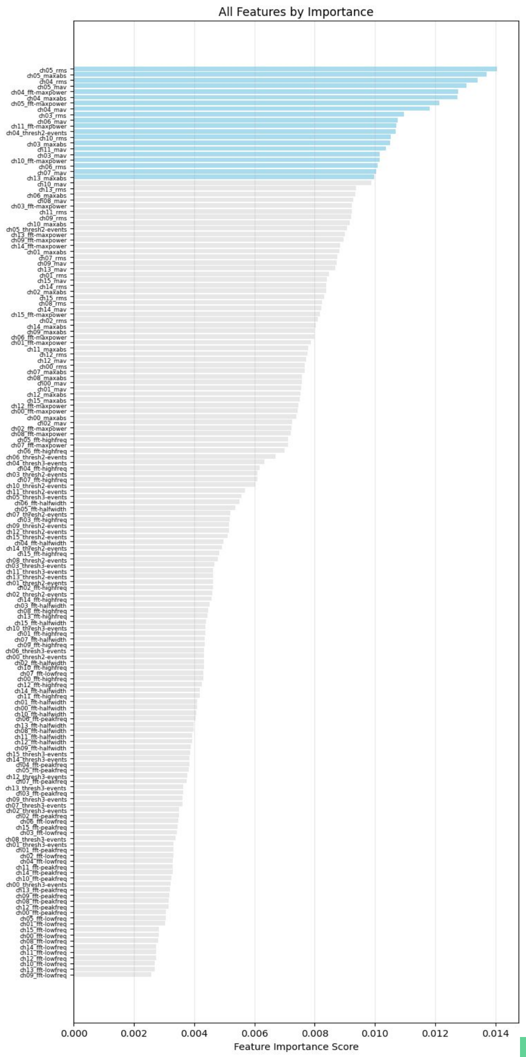
- FFT half width
- FFT highest freq.
- FFT lowest freq.







10 x 16 = 160 features



Feature Selection

Training a random forest classifier, we chose features based on their importance scores.

After removing features with had high correlation scores with each other, we selected 37 final features.

During our model selection process, we decided to try using all features and compared it with our 37 top features.

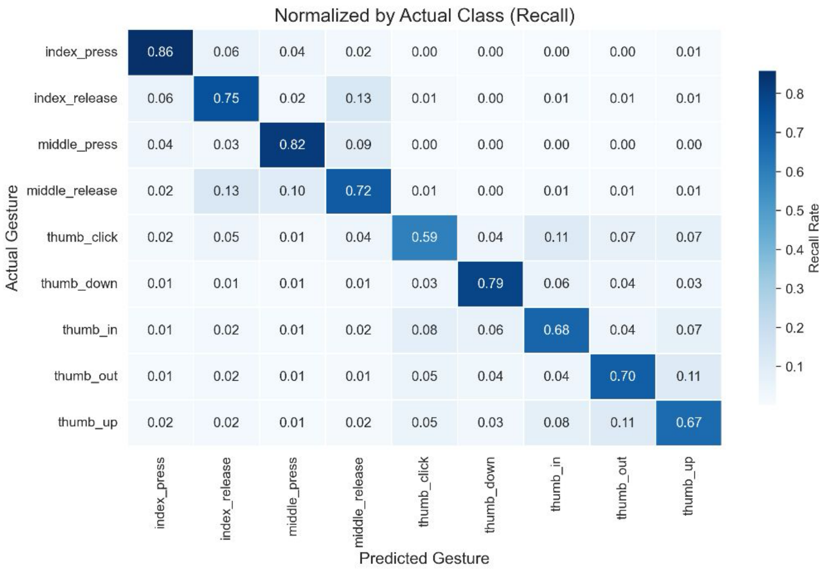
Modeling

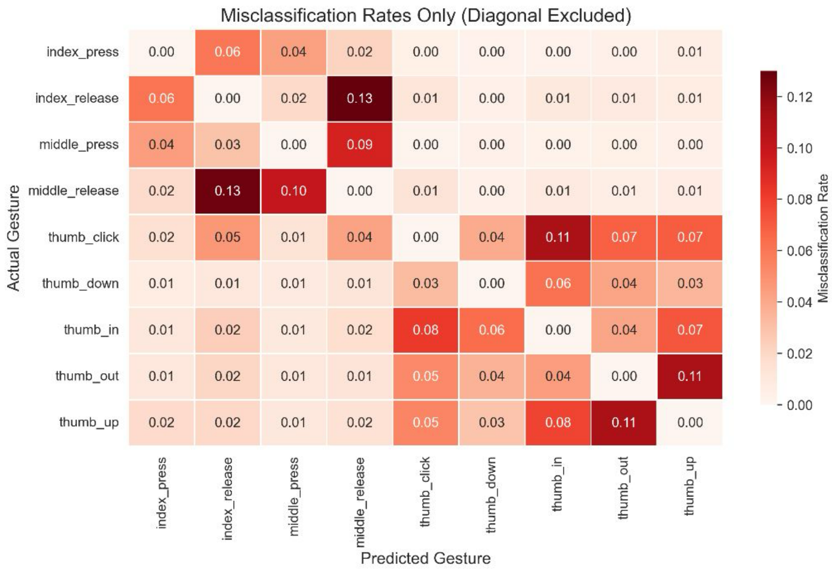
Models trained with 5-fold cross validation::

- Logistic regression with selected features
- Logistic regression (weighted) with selected features
- Logistic regression with all features
- Weighted logistic regression with all features
- XGBoosted Random Forest
- Random forest

The best performer was logistic regression with the selected 37-feature set with a 73.4% accuracy, mean F1 macro 0.72

Confusion Matrix from Logistic Regression Modeling of the 37- feature model



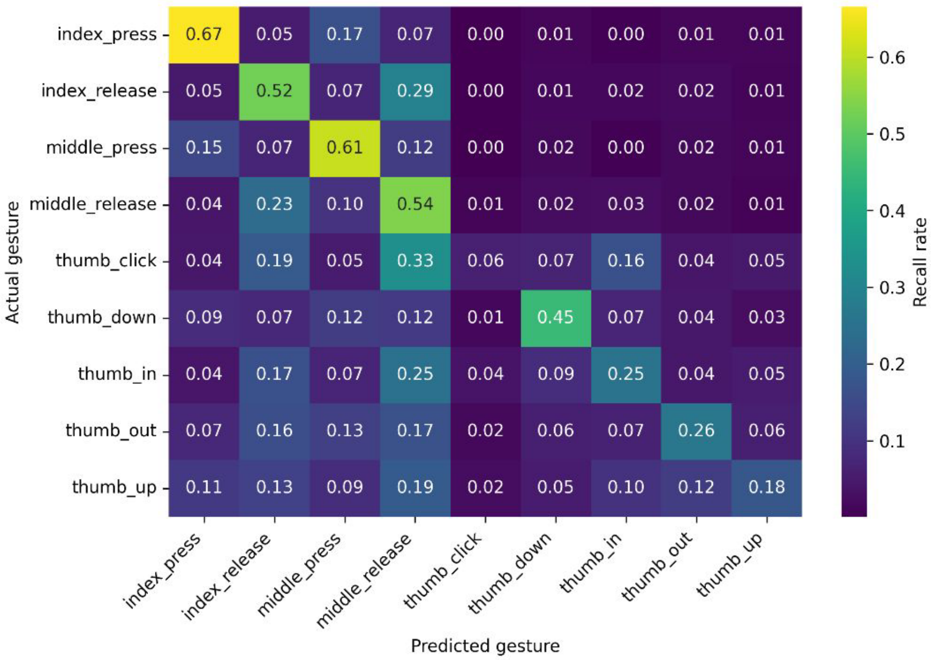


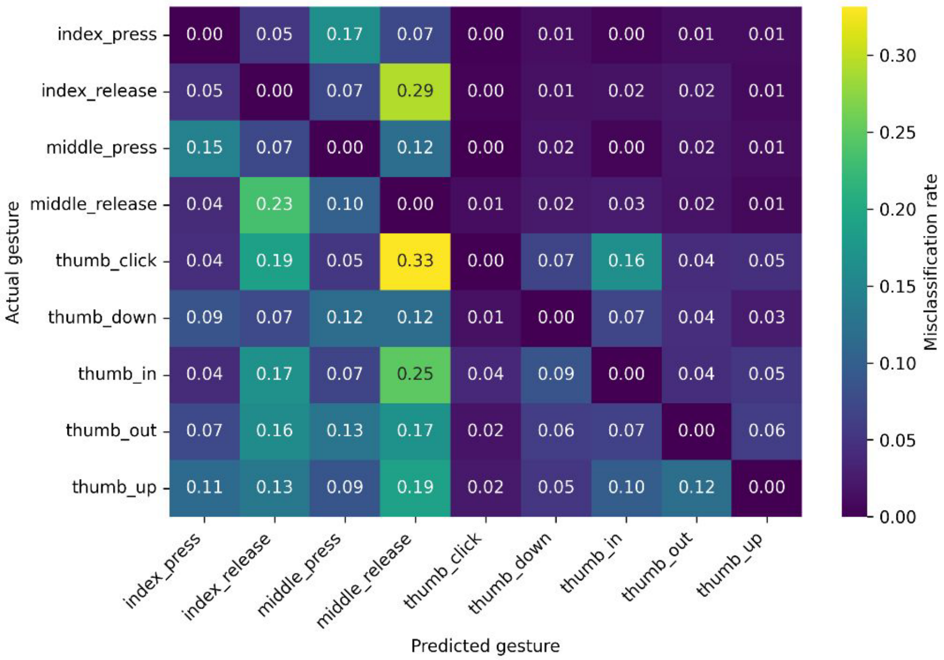
Highest misclassification occurred between either Finger gestures or Thumb gestures

Results

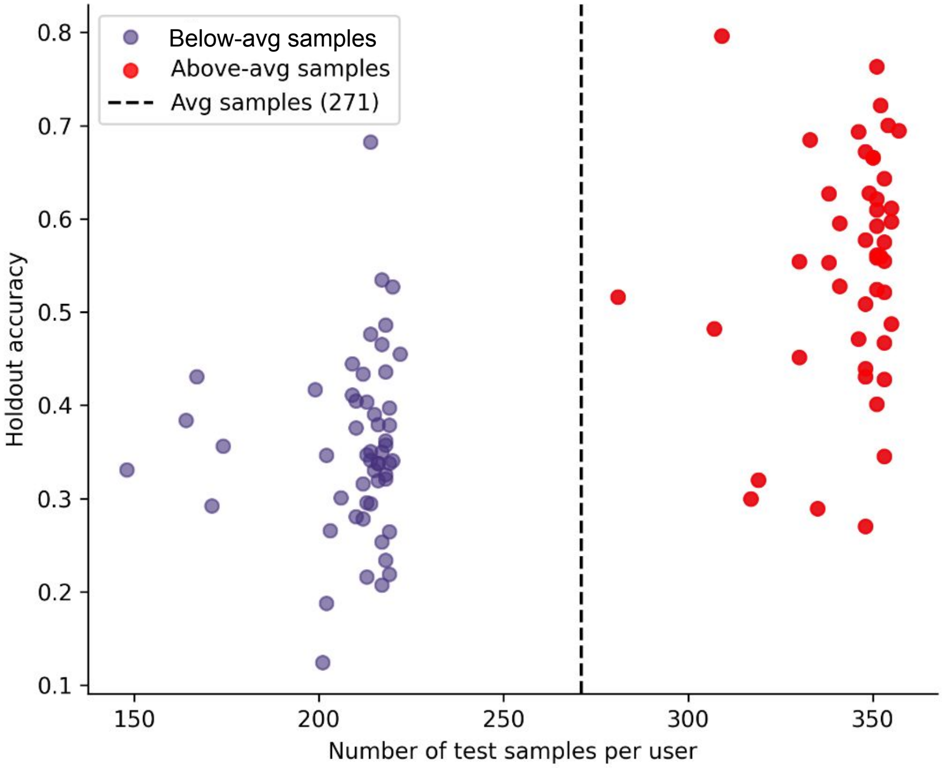
Holdout dataset: 20% gestures from each user

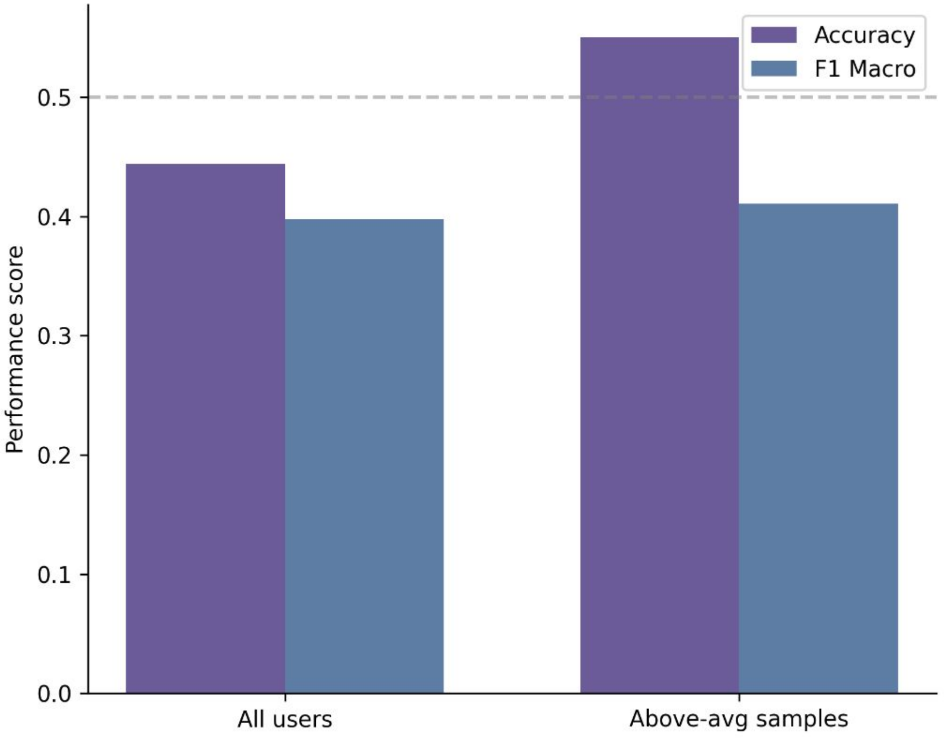
Metric	Value
F1_Macro	0.397683
Accuracy	0.468722
Classification_Error_Rate	0.531278





Results: Those with more trials performed better





Conclusion & Future Directions

- Best model: Logistic regression with L2 regularization on the selected 37-feature set
- Strong CV performance (F1 Macro = 0.72) but considerable generalization gap on holdout data (F1 Macro = 0.40)
- Data quantity matters: Users with more data achieved 15% higher accuracy
- Future improvements
  - Adaptive models: LSTMs, deep learning, template matching ➔ improve generalization
  - Improve trial alignment and time window selection ➔ capture exact muscle movement

Acknowledgments

- Hatice Mutlu
- Steven Gubkin
- Alec Clott
- Erdős Institute