

# Exploring User Accessibility and Human-Machine Interaction Using EMG - Preliminary Report

Isha Arora, Sai Rahul Ponnana, Tanmay Gadgil & Tarandeep Singh

November 3, 2023

**GitHub:** [GitHub Repo](#)

## 1 Summary

The advent of smart devices capable of interpreting and responding to gestures has ushered in a new era of accessibility and independence for those facing mobility challenges. State-of-the-art devices such as gesture-controlled wheelchairs, home automation devices, and prosthetic arms/limbs can be extremely helpful in navigating the day-to-day for the disabled population. Verifying and authenticating the use of gesture-controlled devices for individuals with mobility issues is paramount to ensure their accessibility, effectiveness, and safety.

One way to enable gesture control in these devices is to make use of EMG (electromyography) signals. EMG (electromyography) can be used in wearable devices to detect muscle activation and interpret movement intent. By placing EMG sensors on the skin over targeted muscles such as fingers, or wrists, small electrical signals generated during muscle contractions can be measured. The goal of this project is to use electromyography (EMG) data to build machine-learning models that can accurately classify different hand gestures. The motivation is to explore new applications in human-machine interaction, especially for improving accessibility for people with disabilities.

The dataset contains EMG data collected from 43 participants performing 17 gestures, including 16 distinct gestures like wrist flexion/extension, hand close/open, and finger pinches. The 17th gesture was resting data. The raw EMG signals were pre-processed to extract features like mean absolute value, waveform length, Willison amplitude, etc. that characterize the signals.

Several supervised machine-learning models were trained on the extracted features to classify the gestures, including logistic regression, gradient-boosted trees, support vector machines, and neural networks. The best-performing model was a neural network which achieved 91% accuracy. The models were able to effectively classify most gestures, with the easiest gestures having recall scores around 80-88%, while the hardest gestures had poorer recall around 3-12%.

Overall, the results demonstrate the feasibility of using EMG for gesture recognition. With further optimization, EMG-based gesture interfaces could enable natural and intuitive control mechanisms for improving accessibility.

## 2 Methods

### 2.1 Data Collection and Parsing

The data was sourced from PhysioNet which is a web-based resource designed to support current research and stimulate new investigations in the study of complex physiologic and clinical data. The data can be found at: [GRABMyo](#). [4][6][3]

Data was collected from three distinct sessions, labeled Session 1, Session 2, and Session 3. Each session was stored in separate directories. Within each session, information pertaining to each participant was parsed for subsequent analysis [5]. A comprehensive set of features was extracted from the parsed data. These features encompassed attributes from both the time and frequency domains [1]. The features were calculated for all of the 28 sensors, providing a comprehensive representation of the signal characteristics. The parsed and feature-extracted data from Session 1, Session 2, and Session 3 were integrated into a single cohesive dataset for further analysis.

The data set includes 17 gestures, 16 of which are as listed in [6.1](#). The 17th gesture is the resting data collected after each trial.

## 2.2 Feature Extraction

A critical attribute of any project involving the processing of signal data is to extract meaningful features [2]. We look to generate features by aggregating statistical attributes of the waveforms from the sensor. Our sensor suite has 32 sensors, of which we use 28 sensors that are attached to the wrist and forearm of the user as in ???. Each gesture has data for 5 seconds recorded at a frequency of 2048 Hz. 6.1 shows an example of what a waveform of each gesture trial looks like.

Our first goal is to create reasonable chunks of this 5-second waveform [?]. From these slices, we aim to extract statistics such as mean, max, min, peak, skew, etc. A detailed list of features has been mentioned in the appendix section.

## 2.3 Exploratory Data Analysis

In an attempt to further understand the data and the features extracted, visualizations for each of the features were created.

Visualizations were separated by the electrode ring (4 electrode rings - 2 on the forearm and 2 on the wrist) and line graphs representing average feature value by gesture were created, with each line representing a specific electrode signal (these visualizations can be found in the appendix in 6.5). 16 relevant features were plotted in such fashion and included: Integrated EMG, Mean Absolute Value, Simple Squared Integral, Myopulse Percentage Range, Waveform Length, Differential Value, Differential Absolute Standard Deviation, Willison Amplitude, Power, Peak, Peak-to-Peak, Crest Factor, Skewness, Kurtosis, Form Factor and Pulse Indicator.

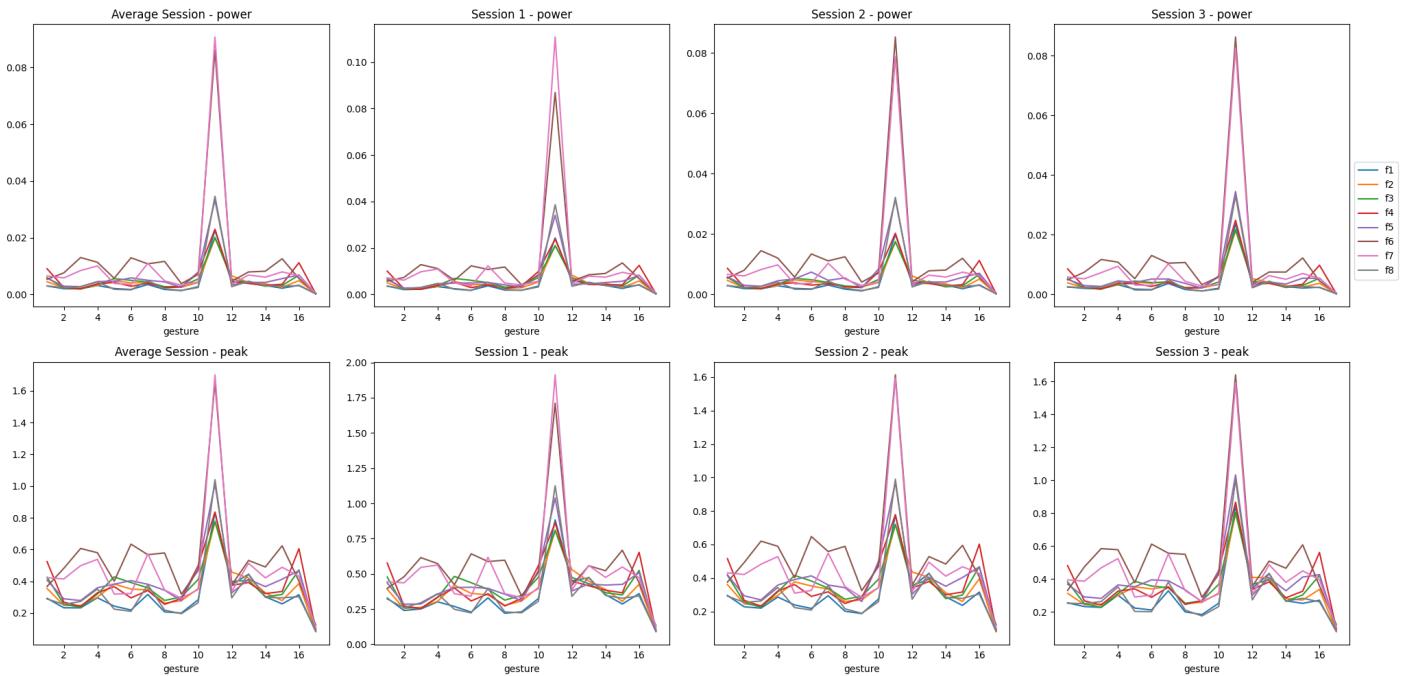


Figure 1: Features for Forearm Ring-1 for each Signal by each Gesture

- While there is some variability, the general profiles and patterns of the features are largely consistent across sessions. The features seem to capture the inherent properties of the gestures that persist through repetitions.

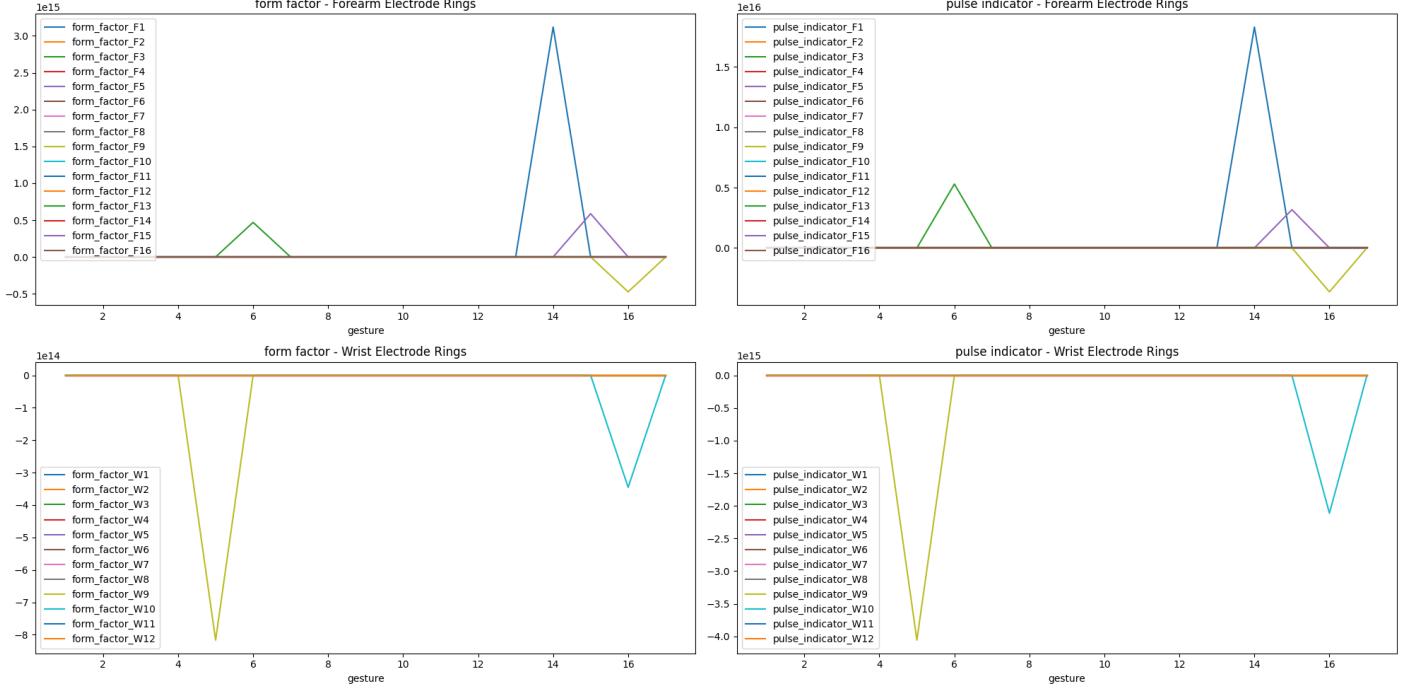


Figure 2: Comparison of Forearm and Wrist Electrode Signals by Gesture for Form Factor and Pulse Indicator Features

- The pulse indicator is higher for the forearm electrodes compared to the wrist. This suggests the forearm electrodes better capture pulse characteristics within the signal. Overall, the wrist electrodes seem more responsive and better at capturing muscle activity, while the forearm ones better capture subtle pulse information. The combination provides complementary information about the gestures performed.

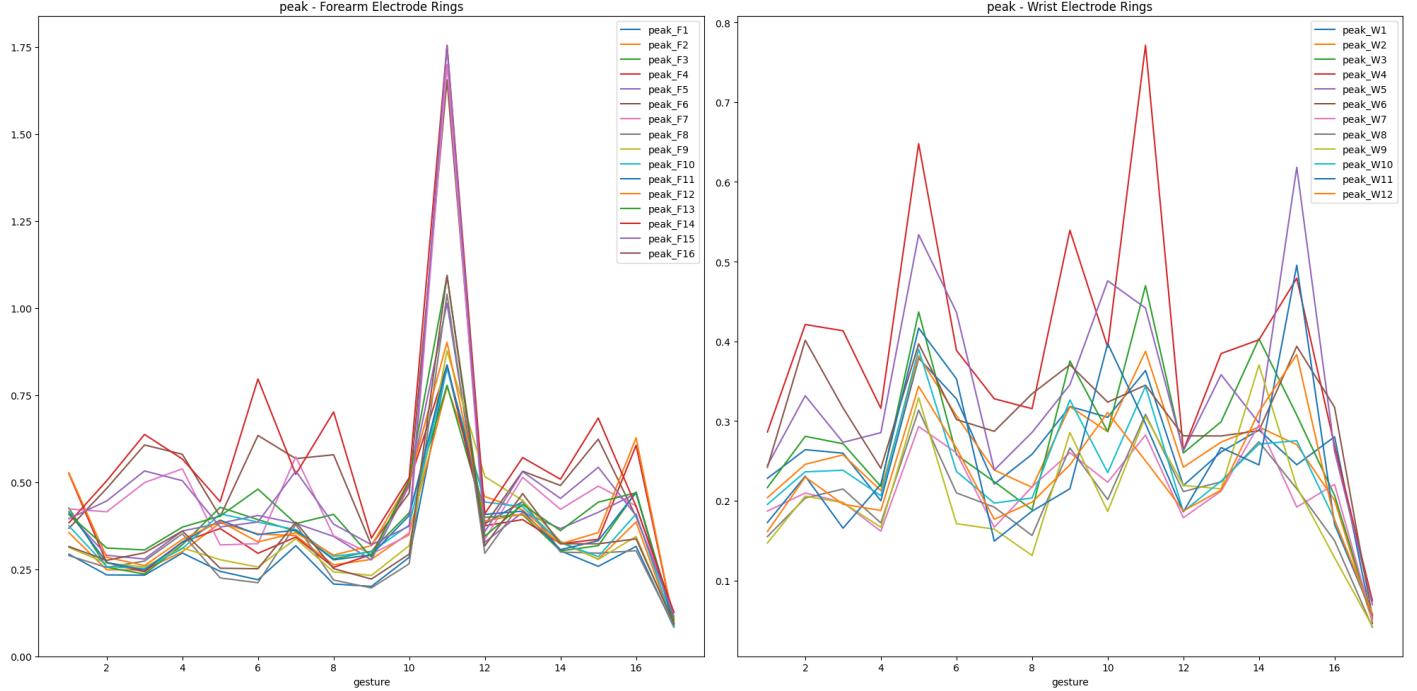


Figure 3: Comparison of Forearm and Wrist Electrode Signals by Gesture for the Peak feature

- An overall trend that was noticed was that signals from the wrist electrodes were comparatively more responsive as

compared to the ones from the forearm for most gestures.

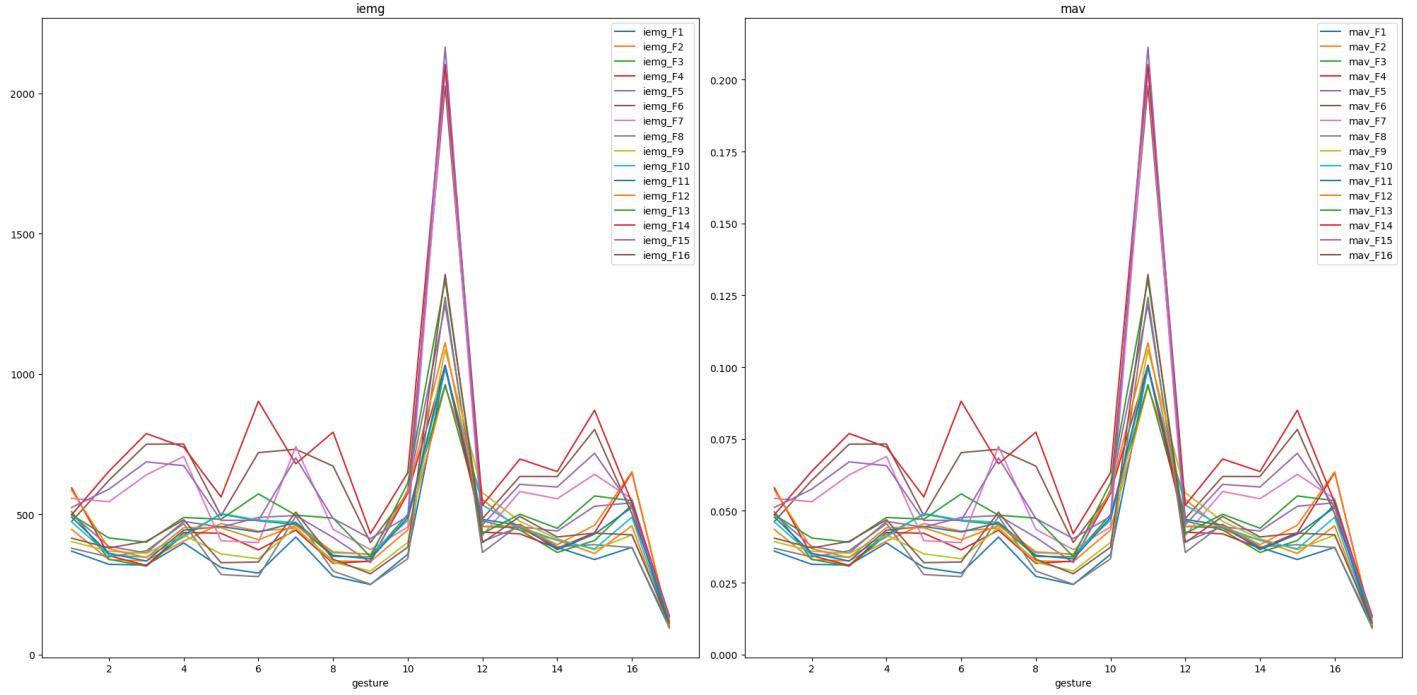


Figure 4: Average Forearm Electrode Values for IEMG and MAV Features for each Signal by each Gesture

- Gestures like **Wrist extension (Gesture 11)** show more distinct patterns compared to other gestures, which show more overlap between gestures and are indicative of more forearm muscle activation as compared to other gestures (as also seen in Figure 5).

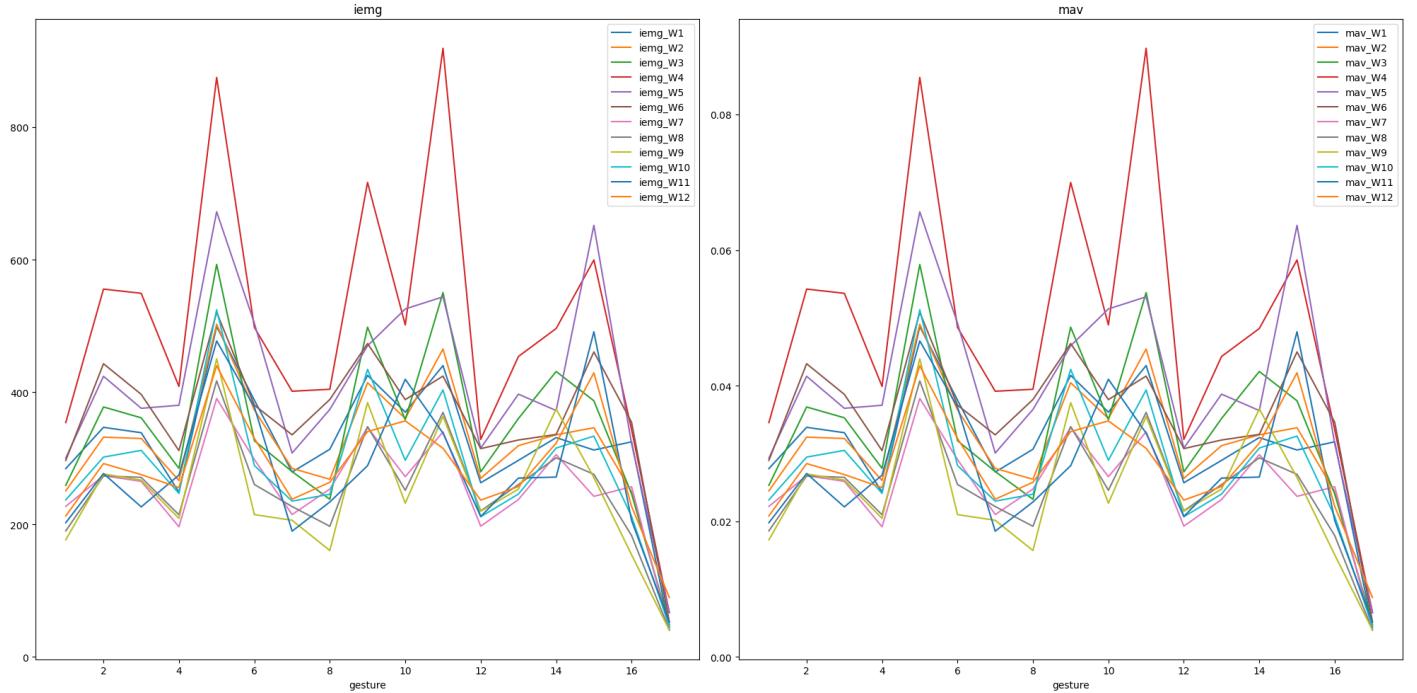


Figure 5: Average Wrist Electrode Values for IEMG and MAV Features for each Signal by each Gesture

- The **red line** signifies the signal from wrist electrode 4 which is situated on the back of the wrist. It is consistently higher for all gestures, over almost all features.
- **Resting data (Gesture 17)** shows consistently low values for most features as expected since there is minimal muscle activity at rest.
- The features exhibit clear differences across the various gestures. This indicates that the features are able to effectively characterize and distinguish between the different gestures.

## 2.4 Modeling

Our main task is to build a classification model to identify a gesture from the features that we have extracted. We also investigated the performance of the model when sliced between participants and sessions.

We have built a suite of models for this task. We start off with a set of simple linear models like Support Vector machines and Logistic Regression as baseline models. After this, we experimented with gradient-boosted trees, Neural networks, and ensembling models. These models were rigorously analyzed on various aspects: performance using accuracy, F1-score, recall, etc.; overfitting via learning curves; confusion matrices to reveal difficult gestures; class activation maps to highlight influential input regions; SHAP values for feature importance; and ablation studies to analyze feature impacts. This comprehensive analysis provided insights into the models' capabilities and limitations, guiding future improvements. Overall, the neural network performed the best. More details can be found here [6.3](#)

Our model analysis involves investigating what features contribute to performance the most. For model training, the extracted EMG signal features were directly fed into the various classification models as a large set of input variables. Rather than applying any feature selection or additional processing, the full collection of features from the feature extraction stage was passed directly as inputs to the models. This provided the algorithms with the complete set of descriptive metrics summarizing the raw EMG signals, allowing the models to determine the optimal features to utilize for gesture classification. By supplying the extensive feature array directly, the models could learn which features were most relevant and influential for distinguishing between gestures based on the underlying relationships learned during training. A variety of supervised machine learning models were trained and thoroughly analyzed for the multi-class gesture classification task:

## 3 Results

### 3.1 Model Performance

Machine Learning Model	Accuracy Score	F1 Score
Neural Network	<b>0.91</b>	<b>0.90</b>
Ensemble Model (LR + RF)	0.89	0.89
Gradient Boosted Tree	0.87	0.87
Logistic Regression	0.88	0.88
Support Vector Machine	0.79	0.79
LSTM	0.91	0.90

Table 1: Machine Learning Model Performance

Model performance is highlighted in table 1. The results show neural networks overall performed the best. The other complex models like gradient-boosted trees and LSTM also performed well. Although the NN was the best-performing model, it did have a high computational demand. The ensemble model provides a good tradeoff between performance and computational cost.

### 3.2 Gesture Analysis

The models were evaluated on their ability to classify each of the 16 distinct gestures with a 17th gesture which represents the rest state. The confusion matrix visualization revealed differences in performance across gestures.

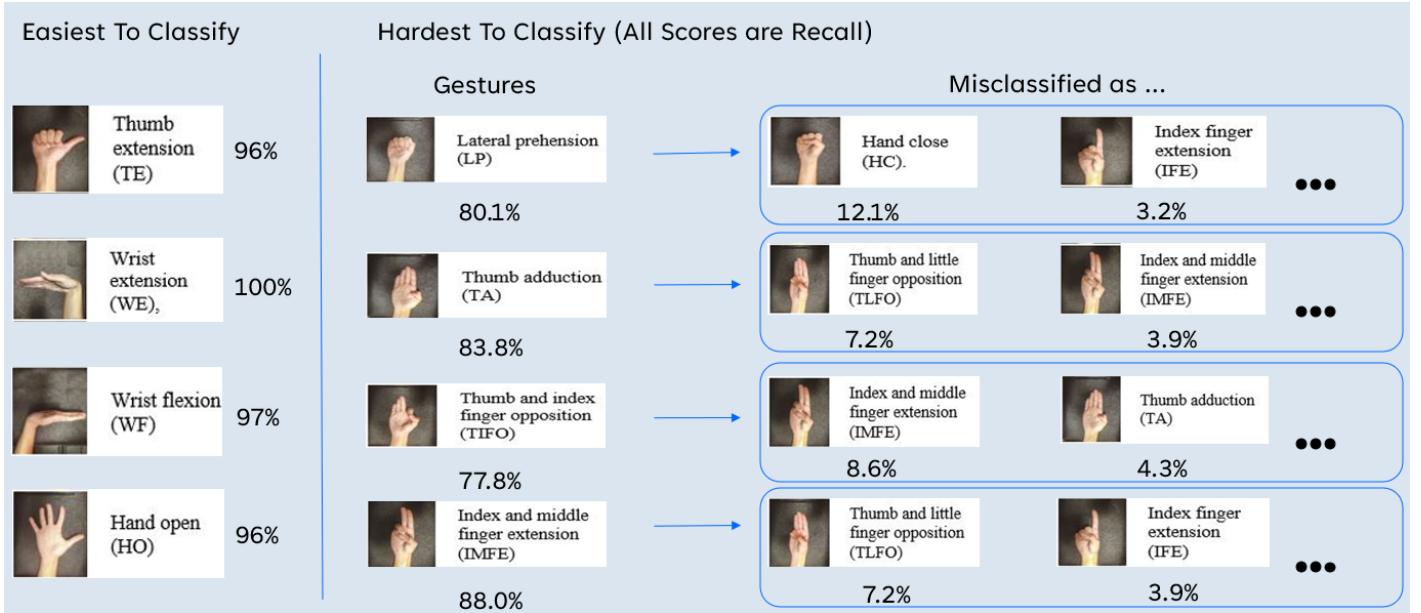


Figure 6: This figure highlights the easiest and hardest gestures for the models to classify. On the left, we have the easiest gestures and on the right, we have gestures that are the hardest to classify. The images on the right are the ones that these gestures are the most confused about.

Certain gestures involving large, distinct movements like wrist flexion and extension were classified very accurately. Wrist flexion (Gesture 11) had a recall score of 97%, while wrist extension (Gesture 12) scored 100%. The substantial wrist flexions and extensions generate clear and consistent muscle activation patterns that are easier for the models to detect.

In contrast, the subtle finger-pinching gestures were the most difficult to classify reliably. The thumb and index finger opposition (Gesture 4) had a recall of just 3.2%, while the thumb and middle finger opposition (Gesture 7) scored only 7.2%. These pinches involve small muscles in the forearm and produce localized, nuanced signals that are harder to differentiate.

The hand close (Gesture 16, recall 12.1%) and thumb and little finger opposition (Gesture 3, recall 7.2%) gestures also posed challenges. There appears to be inherent variability in the forearm muscle patterns during these twisting motions that make them more ambiguous.

This analysis of model performance on individual gestures reveals limitations in detecting certain subtle finger and wrist movements. Additional data collection and model optimization, especially for the pinches, pronation, and supination, could further enhance classification accuracy across all gestures. However, the results demonstrate a solid proof-of-concept for EMG-based gesture recognition.

### 3.3 Feature Analysis

The models were analyzed to determine which EMG signal features contributed most to gesture classification performance.

Feature importance scores were calculated on the Neural Net Model using SHAP values based on how much each feature affected the model output and using the feature importance characteristics of an XGBoost Model. These plots can be found in the appendix here [6.4](#) and [6.4](#). The top features that had the greatest impact were the features that involved differencing the wave with itself. These include Waveform Length (WL), Willison Amplitude (WAMP), DASD, and Integrated EMG (IEMG). Other aspects that directly extracted information from the wave also played an important role like max, min, skew, and peak difference (P2P) These features characterize overall signal intensity and variation, which helps differentiate muscle activations.

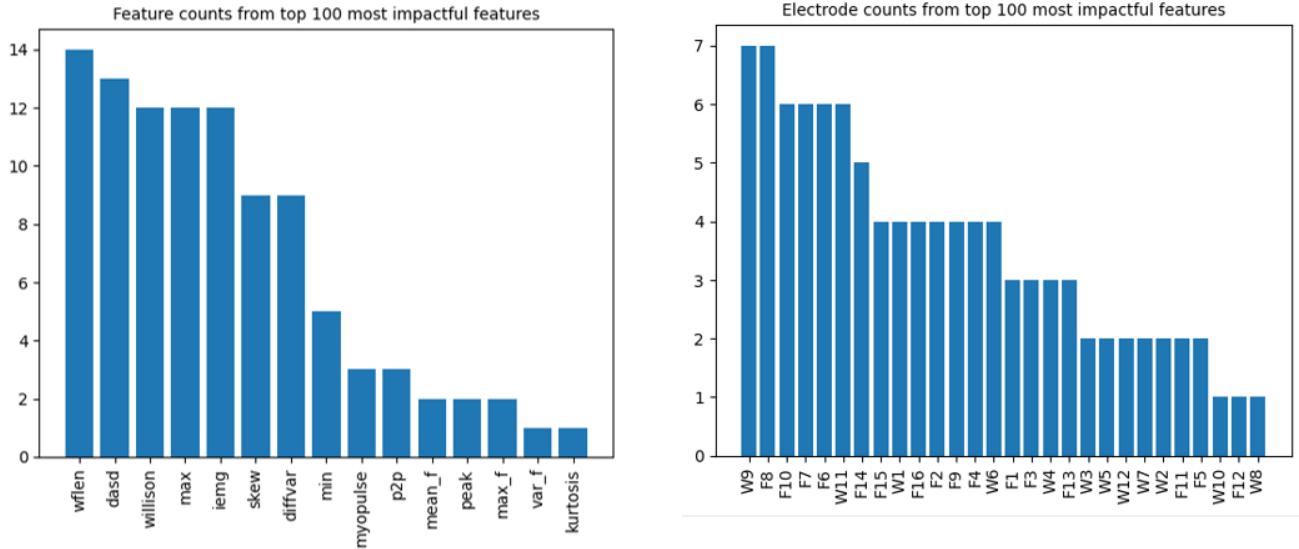


Figure 7: Count of Features and Electrodes from the Top 100 most important features. We consider only the top subset since the entire feature space is  $R^{756}$

We see an equal representation of wrist and forearm electrodes. Some electrodes like W9 and F8 have a stronger impact than the rest but the dropoff in importance is not as significant as the wave features.

### 3.4 Analysis on folds by participant and session

We see an interesting change in the performance characteristics when we create test splits that stratify by participants and testing sessions.

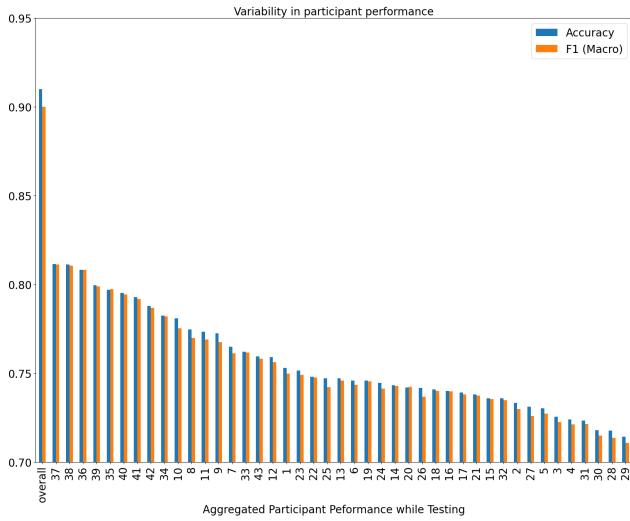


Figure 8: This chart represents the aggregated performance loss when a participant was part of a split. Splits were created by adding participants to testing splits of 5 and the numbers for accuracy were obtained by aggregating by participant

The splits were created in the following fashion - if we are considering participant  $n$  we trained several models where the participant was not in the training set and was only represented in the test set. The test scores for these models were then aggregated to get a value for participant  $n$ . We see a drop in performance in this figure 8 when we stratify by participant. This could be due to a difference in the structure of musculature and the way these muscles activate when a gesture is performed.

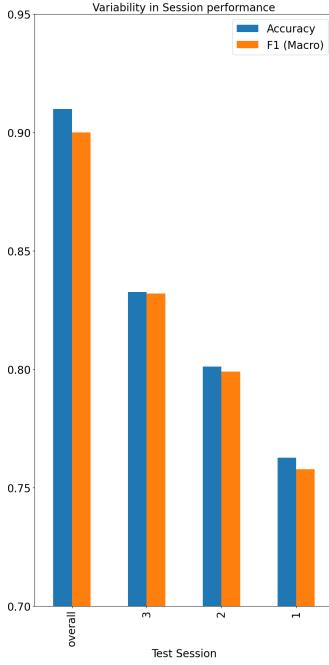


Figure 9: This chart represents the performance loss when splits were created by session. each session here represents a test set while the other two splits were in the training set.

We also see a difference in the session data as seen in this figure 9. This gives us an indication that temporal differences can also affect the activation of muscles.

## 4 Discussion

The neural network model performed the best, indicating the benefits of representation learning from raw features. The results demonstrate the feasibility of EMG-based gesture recognition for human-machine interaction. However, there is room for improvement. Certain gestures like pinches were much harder to classify. Additional data collection and more advanced deep learning models like LSTM networks could help improve performance.

This work helps lay the foundation for developing natural, accessible gesture interfaces. The next steps involve testing impaired subjects, analyzing misclassification costs, and exploring user authentication. With further research, EMG-based interfaces could enable intuitive control and improve the quality of life for people with disabilities.

## 5 Statement of Contributions

1. Feature Extraction: Tanmay Gadgil, Sai Rahul Ponnana
2. Exploratory Data Analysis: Isha Arora
3. Modeling:
  - Logistic Regression: Isha Arora,
  - Support Vector Machine: Tarandeep Singh
  - Semi-Supervised Learning: Tarandeep Singh
  - Ensemble: Sai Rahul Ponnana
  - LSTM: Sai Rahul Ponnana
  - Gradient Boosting Model: Tanmay Gadgil
  - Neural Network: Tanmay Gadgil
4. Model Analysis: Tanmay Gadgil

## References

- [1] Reza Bagherian Azhiri, Mohammad Esmaeili, and Mehrdad Nourani. Emg-based feature extraction and classification for prosthetic hand control. *arXiv preprint arXiv:2107.00733*, 2021.
- [2] Emma Farago, Shrikant Chinchalkar, Daniel J Lizotte, and Ana Luisa Trejos. Development of an emg-based muscle health model for elbow trauma patients. *Sensors*, 19(15):3309, 2019.
- [3] Ary L Goldberger, Luis AN Amaral, Leon Glass, Jeffrey M Hausdorff, Plamen Ch Ivanov, Roger G Mark, Joseph E Mietus, George B Moody, Chung-Kang Peng, and H Eugene Stanley. Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215–e220, 2000.
- [4] N Jiang, A Pradhan, and J He. Gesture recognition and biometrics electromyogram (grabmyo)(version 1.0.1). 2022.
- [5] Kyung Hyun Lee, Ji Young Min, and Sangwon Byun. Electromyogram-based classification of hand and finger gestures using artificial neural networks. *Sensors*, 22(1):225, 2021.
- [6] Ashirbad Pradhan, Jiayuan He, and Ning Jiang. Multi-day dataset of forearm and wrist electromyogram for hand gesture recognition and biometrics. *Sci Data*, 9(733), 2022.

## 6 Appendix

### 6.1 Data Collection and Parsing

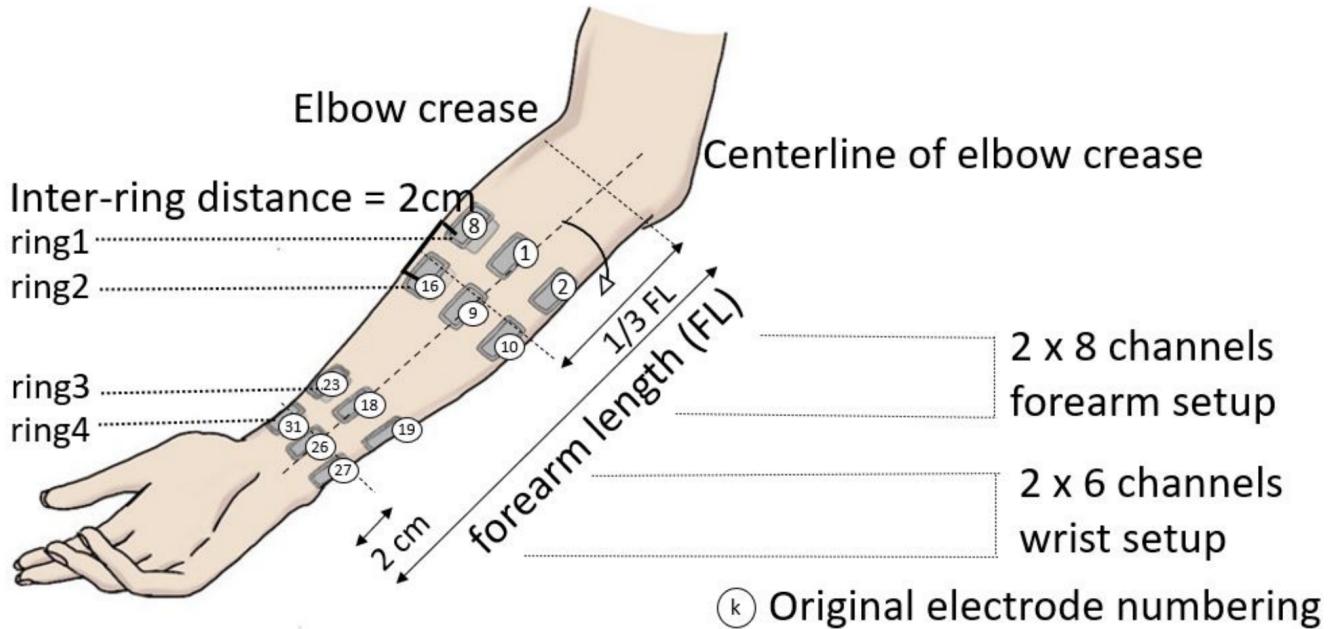


Figure 10: Ring Setup Channels

TABLE GESTURE LIST			
Gesture	Description	Gesture	Description
	Lateral prehension (LP)		Index finger extension (IFE)
	Thumb adduction (TA)		Thumb extension (TE)
	Thumb and little finger opposition (TLFO)		Wrist flexion (WF)
	Thumb and index finger opposition (TIFO)		Wrist extension (WE),
	Thumb and little finger extension (TLFE)		Forearm pronation (FP)
	Thumb and index finger extension (TIFE)		Forearm supination (FS)
	Index and middle finger extension (IMFE)		Hand open (HO)
	Little finger extension (LFE)		Hand close (HC).

Figure 11: Gestures in the Data Set

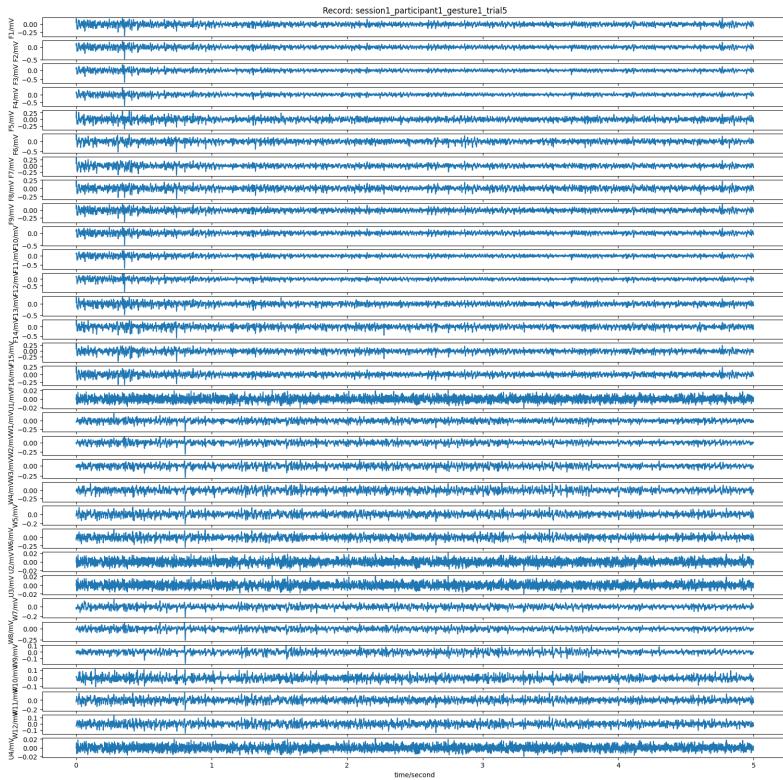


Figure 12: Example Waveform as by Signals

## 6.2 Feature Extraction

1. **Mean:**  $\frac{1}{N} \sum_{i=1}^N x_i$  Useful for assessing overall muscle engagement.
2. **Min:**  $\min(x)$
3. **Max:**  $\max(x)$

Both max and min offer insights into the range of signal amplitudes, indicating the lowest and highest levels of muscle activity observed.

4. **RMS:**  $\sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2}$  Represents the effective amplitude of the signal, providing a measure of its energy content.
5. **Std:**  $\sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \text{Mean})^2}$  Quantifies the variability or dispersion of signal values around the mean, indicating signal stability.
6. **Power:**  $\frac{1}{N} \sum_{i=1}^N x_i^2$  Reflects the total power or energy content of the signal, which is vital for understanding the intensity of muscle activity.
7. **Peak:**  $\max(|x|)$  Identifies the maximum absolute value of the signal, indicating moments of intense muscle contraction.
8. **P2P (Peak-to-Peak):**  $\max(x) - \min(x)$  Provides insights into signal variability.
9. **Crest Factor:**  $\frac{\text{Peak}}{\text{RMS}}$  Indicates how "peaky" or sharp the signal's peaks are relative to its RMS value, which can be relevant in assessing signal shape.
10. **Skew:**  $\frac{1}{N} \sum_{i=1}^N \left( \frac{x_i - \text{Mean}}{\text{Std}} \right)^3$  Skewness indicates asymmetry.
11. **Kurtosis:**  $\frac{1}{N} \sum_{i=1}^N \left( \frac{x_i - \text{Mean}}{\text{Std}} \right)^4$  Kurtosis reflects the tail behavior.
12. **Form Factor:**  $\frac{\text{RMS}}{\text{Mean}}$  Gives information about the shape and distribution of signal amplitudes.
13. **Pulse Indicator:** Provides insights into the measure of pulse characteristics in the signal.
14. **Frequency Domain (DFT):** We apply a DFT on our signal data to convert them into the frequency domain and calculate the following metrics for them: max, sum, mean, var, peak, skew, kurtosis.
15. **IEMG (Integrated EMG):**  $\int_{t_1}^{t_2} |x(t)| dt$  Provides a measure of overall muscle activity.
16. **MAV (Mean Absolute Value):**  $\frac{1}{N} \sum_{i=1}^N |x_i|$  Provides a measure of signal intensity.
17. **SSI (Simple Square Integral):**  $\sum_{i=1}^N x_i^2$  Provides information about the total power or energy content of the signal.
18. **myopulse (Myopulse Percentage Range):** Average of a series of myopulse outputs, and the myopulse output is 1 if the myoelectric signal is greater than a pre-defined threshold.
19. **wflen (Waveform Length):**  $\sum_{i=1}^{N-1} |x_{i+1} - x_i|$  Measures the cumulative length of the signal, reflecting the overall signal amplitude variation.
20. **diffvar (Differential Variance):** Measure of variability in differences between consecutive samples. Provides information about signal stability.
21. **DASD (Differential Absolute Standard Deviation):** Measure of variability in differences between consecutive samples, indicating signal stability.
22. **Willison Amplitude:** Count of times the signal amplitude crosses a threshold, which is important in detecting muscle activation.

## 6.3 Model Descriptions

**Logistic Regression** This simple linear classification model was used as a baseline. While limited in capability, logistic regression is useful for benchmarking.

**Gradient Boosted Decision Trees** We set the classifier and specified 6 CPU threads for parallel processing. Early stopping with a patience of 10 rounds was used to prevent overfitting. Gradient Boosted trees (Especially XGBoost) have been a staple in the data science community. They offer a quick way to build a powerful model that outperforms many other classifiers on tabular data and gives a handy set of tools for model analysis and feature importance. Our model works with 100 boosted estimators with early stopping enabled.

**Ensemble Model** An ensemble model was implemented combining Logistic Regression and Random Forest base models via soft voting. The two complementary models were initialized with hyperparameters like n\_estimators to 100 and max\_iter to 2000. They were then combined using VotingClassifier and soft voting to average predicted probabilities. This stacking ensemble combines multiple well-tuned models to improve prediction accuracy. The ensemble takes advantage of the individual strengths of Logistic Regression and Random Forest.

**Support Vector Machine (SVM)** In addition to the above-mentioned models, we were interested in checking if there exist hyperplanes that can separate data points corresponding to different gestures. Support Vector Classifiers with different kernel types such as RBF, Sigmoid, and Polynomial were implemented on the dataset. We observed the RBF kernel to have the best performance accuracy followed by polynomial and sigmoid kernels respectively.

**Semi-Supervised Learning using DBScan and Random Forest** An interesting approach to this classification task would be to see if there exists inherent groups of data points with a dominant gesture in each cluster. And then train a regular tree-based classifier on this clustered data and labels. However, on training various instances of the clustering algorithm with different hyper-parameters, the clustering algorithm could never fully identify 17 unique clusters. This would lead to a significant amount of misclassification and 0% recall in cases of some classes (gestures). Hence this approach is not included in the modeling analysis and results section.

**Neural Network** A deep feed-forward neural network was implemented, with multiple dense layers between the input and output. All the features that are extracted were given to the input layer with the first hidden layer with 256 units and maps the 64-unit input into a 256-dimensional representation. This larger layer can learn more complex non-linear relationships between the input features. The second hidden layer compresses the representation down to 128 dimensions. Stacked layers allow hierarchical learning. For every layer ReLU activation was used and a dropout regularization randomly drops 30% while training. This model was trained using categorical cross-entropy loss, Adam optimizer with a learning rate of 0.001, and batch size of 128.

**LSTM Network** This network was implemented with the features that were extracted from feature extraction, scaled, and directly passed to input layers, the first LSTM layer with 256 units and the second LSTM layer with 128 units, with a dropout regularization of 20% and sigmoid activation was used after every layer. The output layer was converted to classes of 17, representing each gesture, making the output layer with 17 units. The model was then compiled with categorical loss entropy as the loss function with a batch size of 128.

## 6.4 Feature Importance

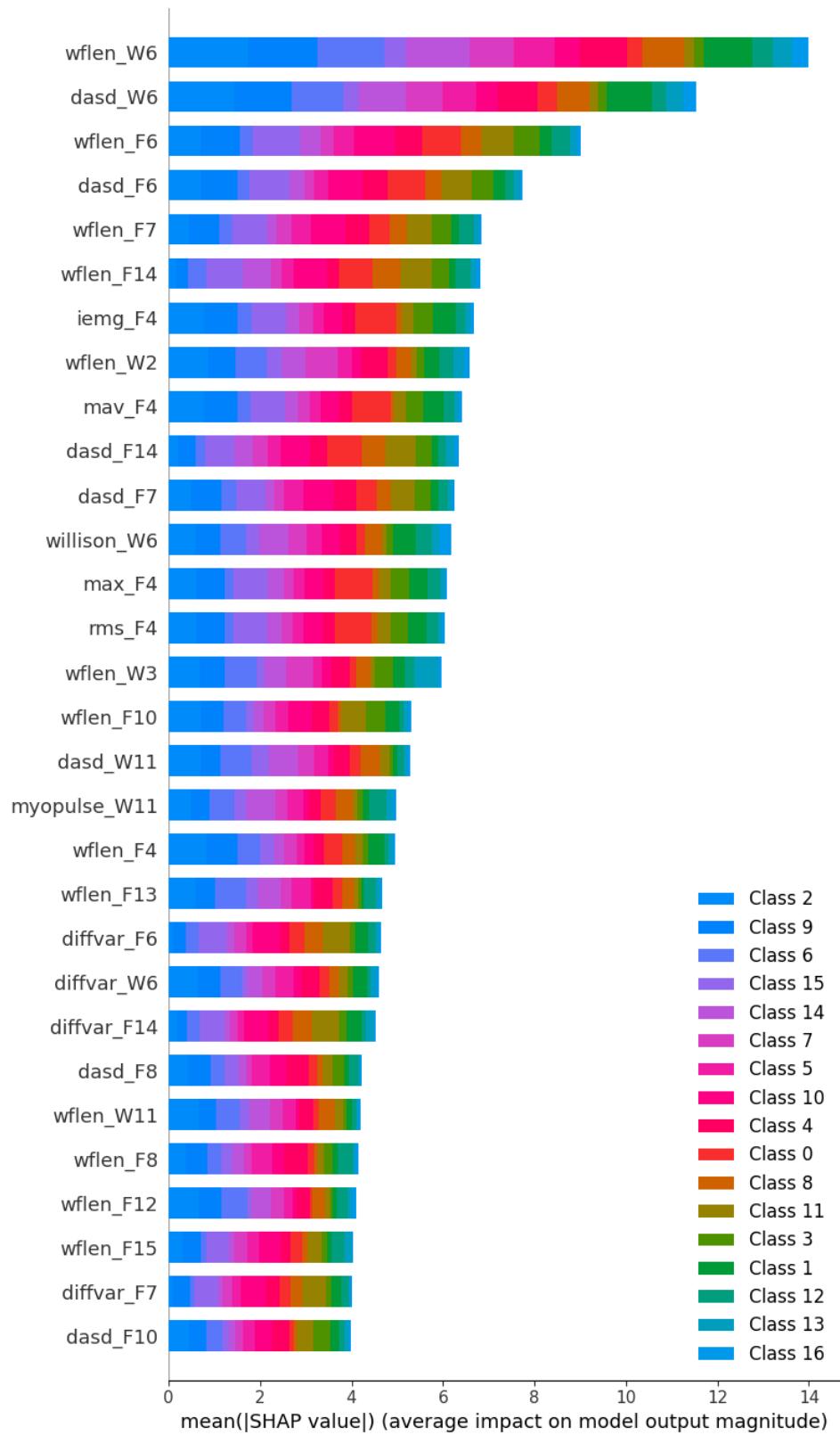


Figure 13: Shapley values for the Neural Network

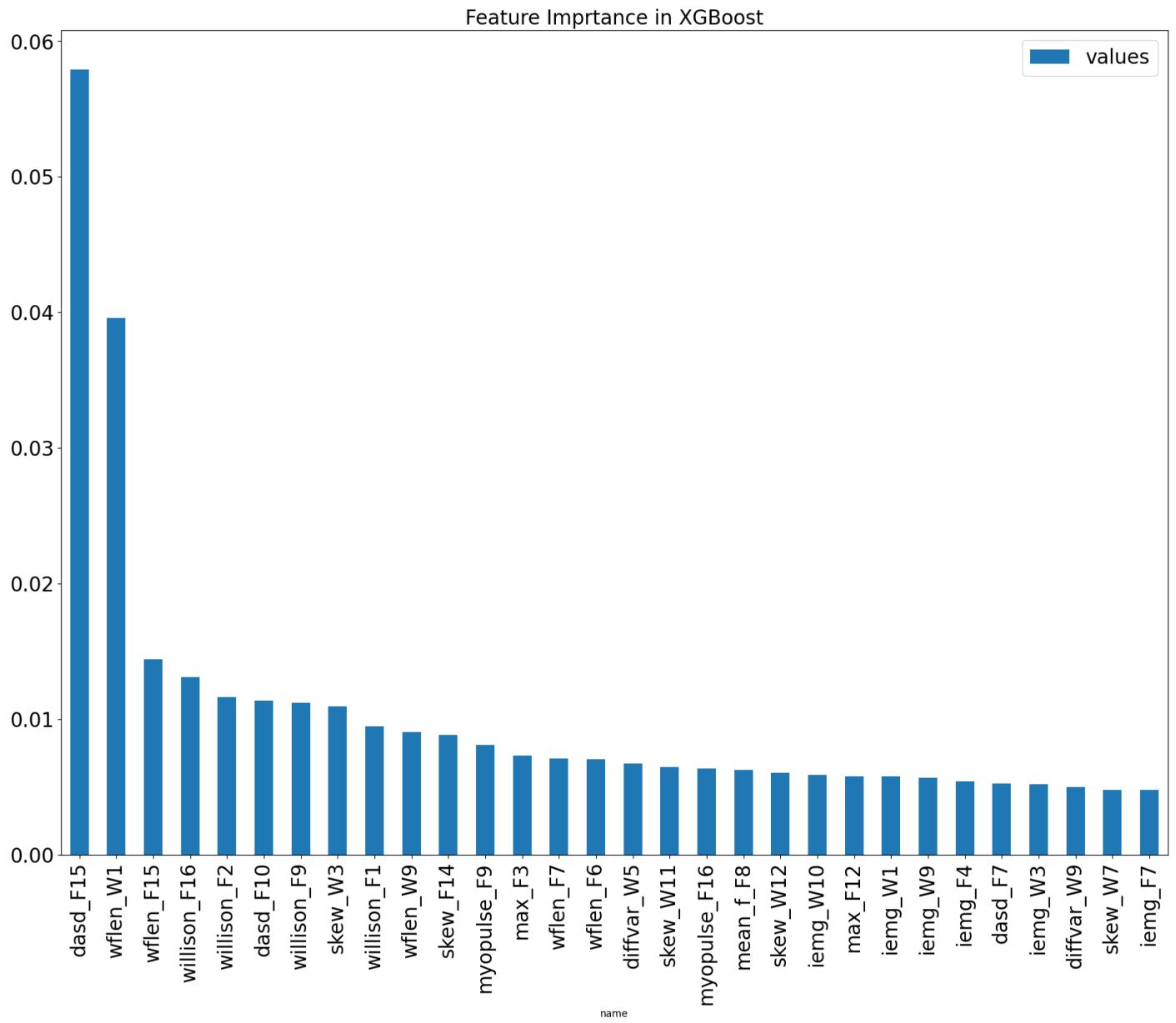


Figure 14: Feature importance values for the XGBoost model

- **Logistic Regression:** A simple linear classification model that achieved 88% accuracy and 0.88 F1 score.
- **Ensemble Model:** The ensemble of gradient-boosted trees, random forest, SVM, and logistic regression achieved 89% accuracy and 0.89 F1 score. Ensembling helps reduce variance and bias.
- **Gradient Boosted Trees:** This model uses an ensemble of weak decision tree models. It reached 87% accuracy and 0.87 F1 score. The trees model nonlinear relationships well.
- **Support Vector Machine (SVM):** The SVM classifier obtained 79% accuracy and 0.79 F1 score. SVMs maximize the margin between classes.
- **Neural Network:** It achieved the best accuracy of 91% and F1 score of 0.90. The neural network's ability to learn complex representations made it very effective for this problem.
- **LSTM:** This recurrent neural network suited for sequence data got 91% accuracy and 0.90 F1 score. It can model temporal dynamics.

## 6.5 Further Visualizations

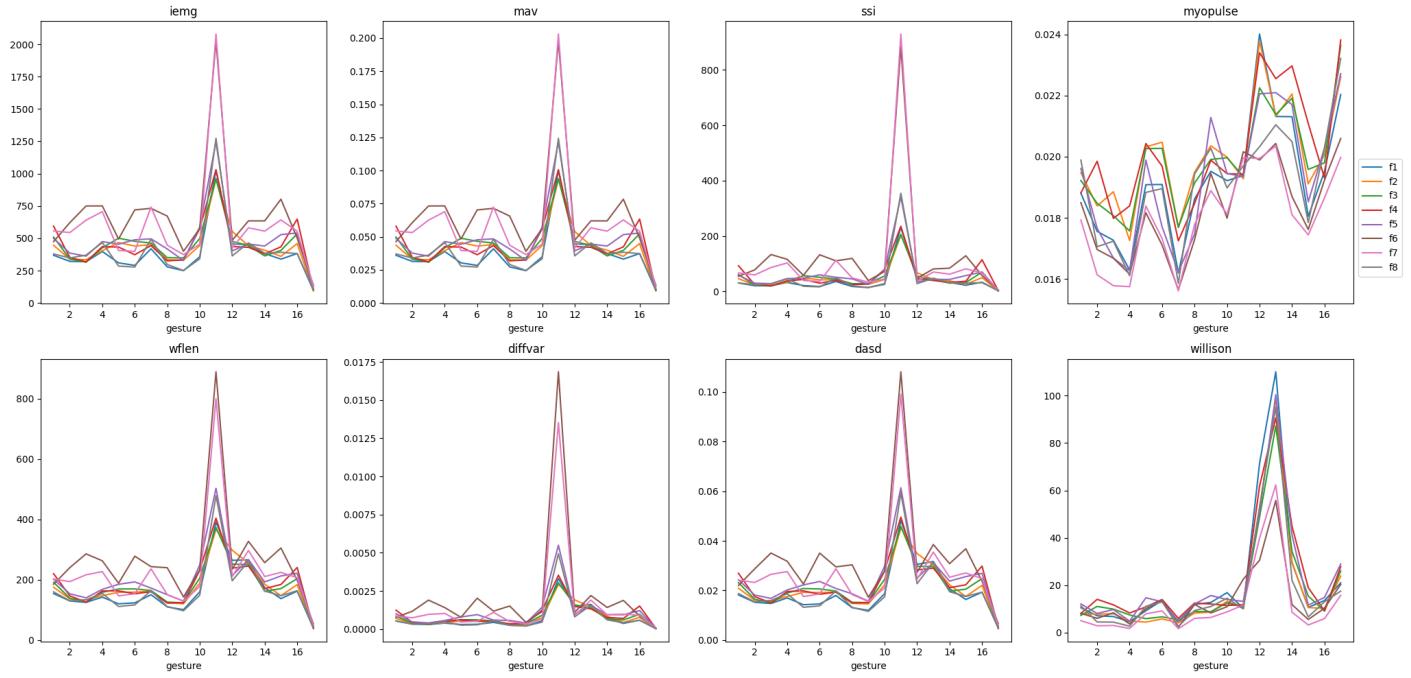


Figure 15: Average Forearm Ring 1 Features for each Signal by each Gesture for first 8 features

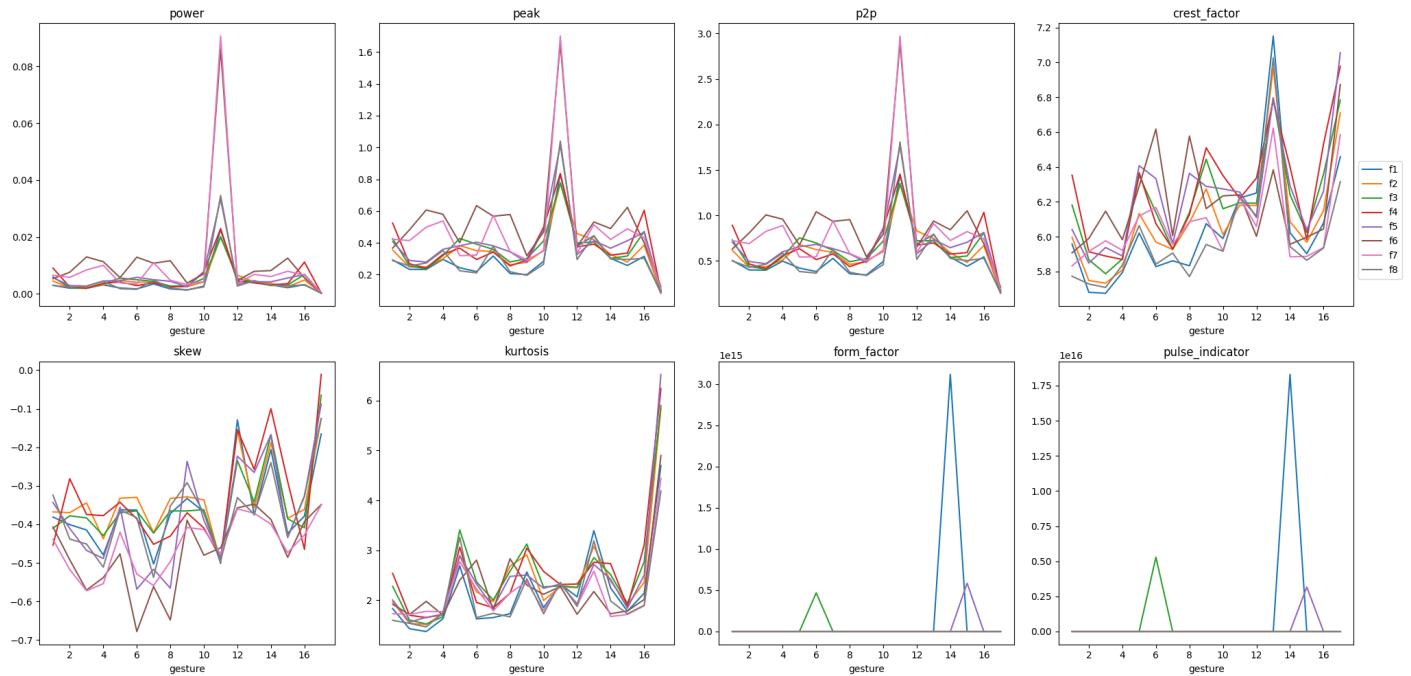


Figure 16: Average Forearm Ring 1 Features for each Signal by each Gesture for last 8 features

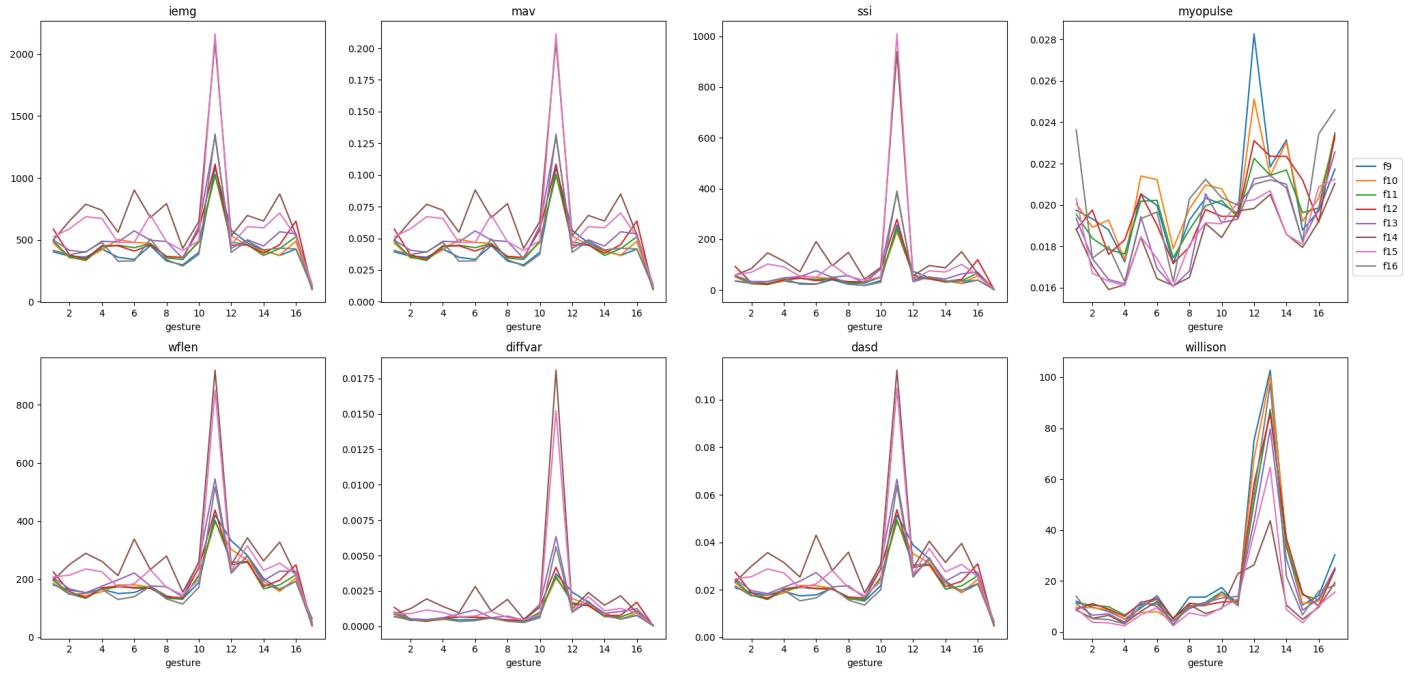


Figure 17: Average Forearm Ring 2 Features for each Signal by each Gesture for first 8 features

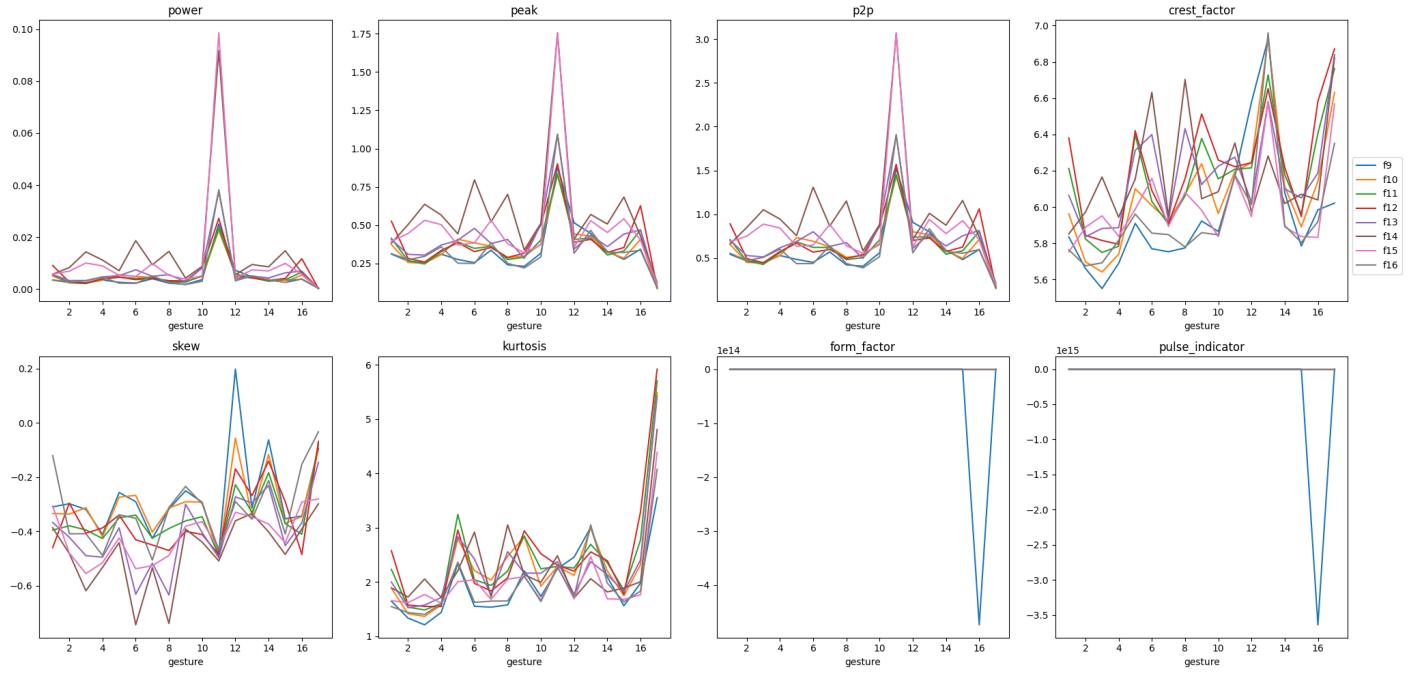


Figure 18: Average Forearm Ring 2 Features for each Signal by each Gesture for last 8 features

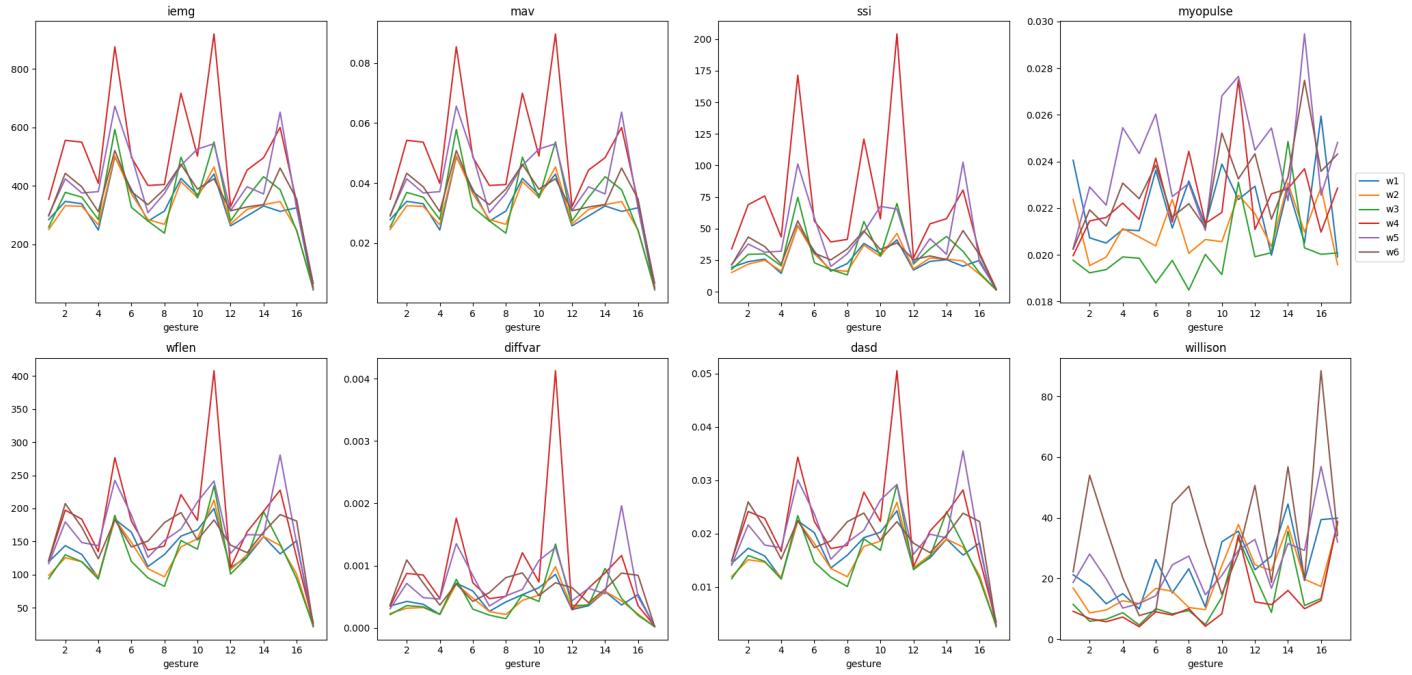


Figure 19: Average Wrist Ring 3 Features for each Signal by each Gesture for first 8 features

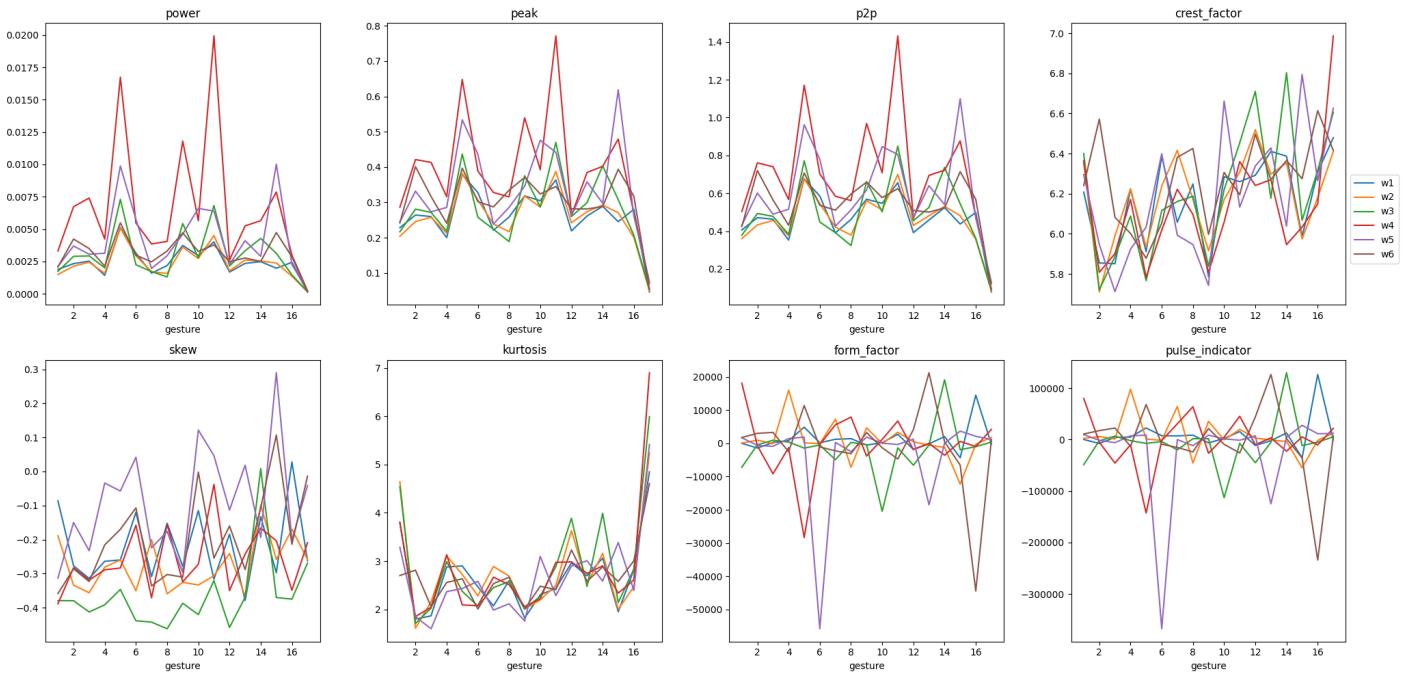


Figure 20: Average Wrist Ring 3 Features for each Signal by each Gesture for last 8 features

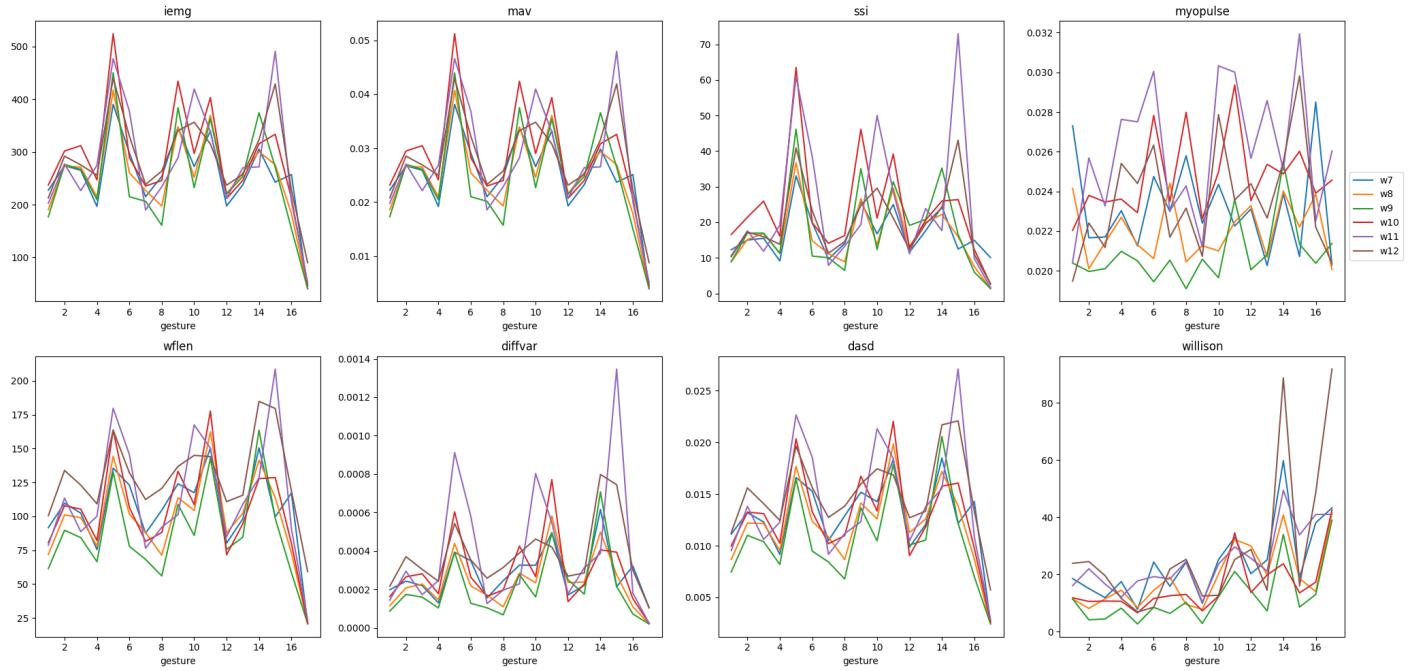


Figure 21: Average Wrist Ring 4 Features for each Signal by each Gesture for first 8 features

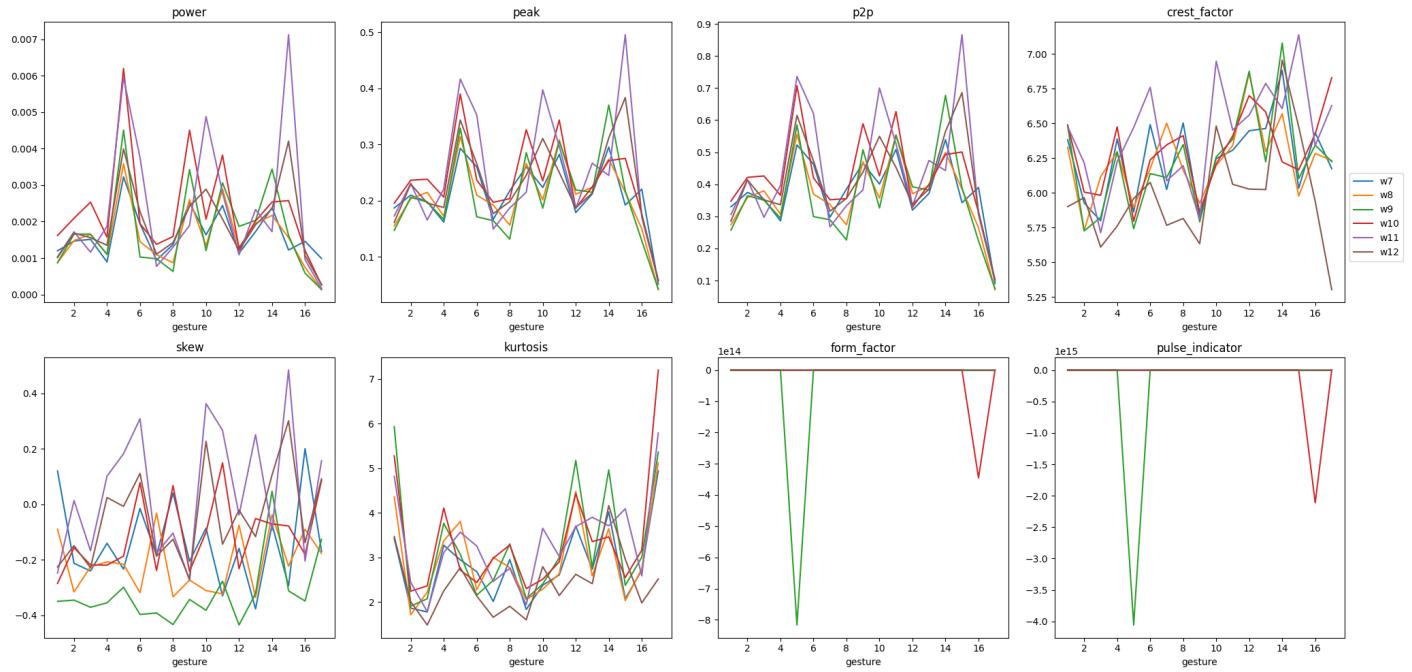


Figure 22: Average Wrist Ring 4 Features for each Signal by each Gesture for last 8 features