Robust sEMG-Based Gesture Classification for the Synapse Gesture Recognition Challenge

Team AtenRise

January 2026

Abstract

This report presents a complete and deployment-ready pipeline for classifying five hand gestures from surface electromyography (sEMG) signals collected from eight forearm channels. The system combines principled signal processing with a compact one-dimensional convolutional neural network (1D CNN) tailored for short temporal windows. The pipeline includes band-pass filtering, sliding-window segmentation, and channel-wise normalization, followed by a three-block temporal CNN trained end-to-end.

Training is performed *exclusively* on the provided Synapse training dataset using a subject-wise split, so that evaluation reflects true cross-subject generalization and avoids any form of data leakage. On a held-out subject, the best model achieves a macro F1-score of approximately 0.70 with comparable accuracy, using a network with well under one million parameters. This strikes a deliberate balance between performance on unseen data, model complexity, and real-time deployability, in line with the challenge evaluation criteria.

1 Introduction

Surface electromyography (sEMG) provides a window into muscle activation and is a key modality for gesture-controlled interfaces, assistive technologies, and human–robot interaction. In this challenge, the task is to convert short, fixed-length sEMG windows from eight channels into one of five gesture labels, using only the provided training dataset.

The evaluation emphasizes three aspects: performance on an unseen test set (accuracy and F1-score), model complexity, and quality of the signal interpretation and methodology. This report therefore focuses not only on the final scores, but on the rationale behind each design choice, and on explicit safeguards against data leakage.

The proposed solution is based on three principles:

1. Respect the physics and statistics of sEMG through appropriate filtering and temporal windowing.

2. Normalize away subject- and session-specific effects to focus on gesture-related structure.

3. Use a compact yet expressive 1D CNN that generalizes well to unseen subjects while remaining efficient.

2 Problem Formulation

Let $x_c(t)$ denote the raw sEMG signal from channel $c \in \{1, \ldots, 8\}$ at discrete time index t. The objective is to learn a function

$$f : \mathbb{R}^{8\times T} \to \{1, \ldots, 5\}$$

that maps a short temporal window of length T samples (here approximately 200 ms) from all eight channels to one of five gesture classes.

Each training example consists of an input window $X_i \in \mathbb{R}^{8\times T}$ and a ground-truth label $y_i \in \{1, \ldots, 5\}$. At inference time, f is applied to consecutive windows extracted from a continuous stream of sEMG.

3 Dataset and Signal Characteristics

The Synapse dataset contains sEMG recordings from 25 subjects performing five different hand gestures across multiple days and sessions. For each session and subject, CSV files store the raw sEMG from eight electrodes placed on the forearm. The sampling rate is approximately 1000 Hz.

From a signal perspective, sEMG is:

- **Stochastic and non-stationary**: amplitude and dominant frequencies change as muscles engage and relax.
- **Band-limited**: informative content lies mostly in the 20–450 Hz range.
- **Multi-channel**: spatial differences across the eight channels reveal which muscle groups are active.

Different gestures produce distinct spatio-temporal activation patterns. A robust classifier must therefore leverage both temporal structure and inter-channel relationships.

4 Signal Interpretation and Processing

4.1 Band-Pass Filtering

Raw sEMG contains slow drifts, motion artifacts, and high-frequency interference. To focus on physiologically relevant content, each channel is processed with a fourth-order Butterworth band-pass filter with passband 20–450 Hz. For each channel c, the filtered signal $\tilde{x}_c(t)$ is obtained by convolving $x_c(t)$ with the band-pass filter impulse response.

This step removes components outside the typical sEMG band, making the subsequent learning problem better conditioned and less sensitive to sensor noise and environmental artifacts.

4.2 Sliding-Window Segmentation

The continuous filtered signals are segmented into overlapping windows. Each window spans 200 ms with a stride of 50 ms, corresponding to 200 samples per window at a 1000 Hz sampling rate and 75% overlap.

For window index k and channel c, the windowed signal is

$$w_{c,k}[n] = \tilde{x}_c(n + kS), \quad n = 0, \ldots, L-1,$$

with window length $L = 200$ and stride $S = 50$. Each window forms a tensor $W_k \in \mathbb{R}^{8\times 200}$.

This sliding-window strategy has two advantages: it increases the effective number of labeled examples, and it enables low-latency predictions as the user transitions between gestures.

4.3 Channel-Wise Normalization

sEMG amplitude varies strongly across subjects, sessions, and electrode placements. To reduce this variability, channel-wise z-score normalization is applied. For each channel c, the mean μ_c and standard deviation σ_c are computed *only on the training windows*:

$$\mu_c = \mathbb{E}[w_{c,k}], \quad \sigma_c = \sqrt{\mathbb{E}[(w_{c,k} - \mu_c)^2]}.$$

Each window is then normalized as

$$\hat{w}_{c,k}[n] = \frac{w_{c,k}[n] - \mu_c}{\sigma_c}.$$

Using training-only statistics is a deliberate choice to avoid data leakage. Validation and test windows are transformed with these fixed parameters, emulating a real deployment where only past data are available to calibrate the system.

5 Model Architecture and Complexity

5.1 Design Rationale

The model must discover discriminative patterns in short, noisy time series while remaining compact. A one-dimensional convolutional neural network (1D CNN) is a natural fit because it:

- Learns temporal filters that resemble classic EMG features (envelopes, onset patterns, frequency bands).
- Combines information across channels to capture spatial activation patterns.
- Scales well to real-time applications and embedded hardware.

The architecture is intentionally shallow and narrow compared to typical deep vision models, to keep the number of parameters low and reduce overfitting risk.

5.2 Network Topology

Each input window has shape (C, T) with $C = 8$ channels and $T = 200$ time steps. The CNN consists of three convolutional blocks followed by global average pooling and a linear classifier:

- **Block 1**: Conv1D with 32 filters and kernel size 7, followed by batch normalization, ReLU activation, and max pooling.
- **Block 2**: Conv1D with 64 filters and kernel size 5, followed by batch normalization, ReLU activation, max pooling, and dropout with rate 0.3.
- **Block 3**: Conv1D with 128 filters and kernel size 3, followed by batch normalization, ReLU activation, and global average pooling over time.
- **Classifier**: fully connected layer mapping the resulting 128-dimensional representation to 5 gesture logits.

This topology yields a parameter count on the order of 10^5 (well below one million parameters). This low complexity directly addresses the challenge criterion on model size while still achieving strong performance.

6 Training Strategy and Metrics

6.1 Subject-Wise Train–Validation Split

To approximate the challenge scenario where the final test set consists of unseen subjects, a subject-wise split is used. All windows from one subject are held out for validation, and all windows from the remaining subjects are used for training.

This design explicitly avoids mixing windows from the same subject across training and validation, which would artificially inflate performance and constitute a subtle form of data leakage.

6.2 Optimization Setup

The CNN is trained with the Adam optimizer using a learning rate of 10^{-3} and weight decay of 10^{-4}. The batch size is 128 windows. The loss function is the categorical cross-entropy between the predicted probabilities and the true gesture labels.

Training is run for up to 50 epochs. After each epoch, the model is evaluated on the held-out subject, and the following metrics are recorded:

- Accuracy.
- Macro-averaged precision, recall, and F1-score over the five gesture classes.

The model achieving the best validation macro F1-score is saved and used for all final inference. This choice reflects the challenge emphasis on F1-score while still monitoring accuracy.

7 Results on Held-Out Subject

On the held-out validation subject, the best model achieves a macro F1-score of approximately 0.70, with validation accuracy of a similar magnitude. These results indicate that the model has learned gesture-specific patterns that generalize beyond the individuals used for training.

The learning curves show rapid improvement during the first few epochs, followed by a plateau with moderate fluctuations. This behavior is typical of a model that has reached a good balance between underfitting and overfitting given the current architecture and preprocessing pipeline.

It is important to note that these validation results are obtained under a stricter subject-wise protocol and serve as a proxy for performance on the organizer's hidden test set. No hyperparameters have been tuned on any form of test data, and no external datasets have been introduced.

8 Alignment with Evaluation Criteria

8.1 Performance on Unseen Data

The subject-wise split is explicitly designed to mimic the hidden test evaluation: the model is trained on a subset of subjects and evaluated on a subject not seen during training. This encourages learning generalizable sEMG patterns rather than memorizing subject-specific signatures.

By monitoring both accuracy and macro F1-score on this held-out subject and selecting the model with the highest macro F1, the training process is directly aligned with the challenge's primary ranking metrics.

8.2 Accuracy and F1-Score

Macro F1-score is chosen as the primary selection metric to ensure balanced performance across all five gestures, rather than optimizing only for overall accuracy. Accuracy is still tracked to quantify the overall fraction of correctly classified windows.

This dual focus provides a strong basis for expected performance on the hidden test set, where class frequencies and subject characteristics may differ from the training data.

8.3 Model Complexity and Parameters

The chosen 1D CNN has a parameter count on the order of 10^5, significantly smaller than typical deep models used in vision or audio. This low complexity:

- Reduces overfitting risk on the available dataset.
- Enables efficient inference on modest hardware.
- Directly addresses the evaluation criterion on model complexity and parameter count.

Despite its compact size, the model delivers robust F1-scores on unseen subjects, demonstrating that careful architecture design can achieve a favorable accuracy–complexity trade-off.

8.4 Technical Implementation vs Methodology

The technical implementation is kept modular and transparent: separate components handle dataset loading, preprocessing, model definition, training, evaluation, and inference. This structure mirrors the conceptual pipeline presented in this report.

At the same time, each methodological choice—filtering band, window size and stride, normalization scheme, network depth and width, subject-wise split, and model selection rule—is motivated by the properties of sEMG and the challenge objectives. Together, the code and this report provide a coherent and reproducible solution.

9 Data Integrity and Leakage Prevention

Given the explicit warning that any data leakage or use of unauthorized datasets will result in disqualification, particular care has been taken to avoid such issues:

- Only the official training portion of the Synapse dataset is used. No external data or pretraining is introduced.
- The subject-wise split ensures that no windows from the validation subject appear in the training set.
- Normalization statistics are computed strictly on the training windows and then reused for validation and inference.
- Model selection is based solely on validation metrics; the hidden test set is never accessed during development.

These safeguards ensure that the reported performance is an honest estimate of how the model is expected to behave on the organizers' hidden test dataset.

10 Conclusion

This report has presented a complete solution for the Synapse sEMG gesture classification challenge, combining a physically informed signal processing pipeline with a compact 1D CNN classifier. The design directly targets the evaluation criteria by emphasizing cross-subject generalization, balanced accuracy and F1-score, low model complexity, and strict avoidance of data leakage.

Future work could explore attention-based temporal pooling, domain adaptation across subjects, and confidence-aware decision strategies, but the current system already provides a strong, principled baseline for robust sEMG-based gesture recognition.