

Synapse — Analytical Report: Signals, Models, and Methods

Robust Subject-Independent 8-Channel sEMG Gesture Recognition

Keerthi Kumar K J

Abstract

This report presents a subject-independent framework for classifying gestures from 8-channel surface electromyography (sEMG) signals. We focus on interpreting the received biosignals, motivating each signal processing step based on observed signal properties, and justifying the selected machine learning architectures through physiological and statistical reasoning. A leakage-safe preprocessing pipeline, robust normalization, and a soft-voting ensemble of complementary neural models are employed. Emphasis is placed on the rationale behind every methodological choice and on the inferences drawn from both processed signals and model outputs, in alignment with real-world deployment constraints.

Contents

1 Signal Interpretation and Physiological Meaning	3
1.1 Nature of surface EMG signals	3
1.2 Exploratory signal evidence	3
2 Signal Processing Methods and Rationale	4
2.1 Filtering and detrending	4
2.2 Windowing strategy	4
2.3 Robust normalization using Median–IQR scaling	4
2.4 Data augmentation	5
3 Machine Learning Approach and Model Architecture	5
3.1 Validation protocol	5
3.2 Model architectures and design rationale	5
3.2.1 NeuroCNN	5
3.2.2 NeuroResNet	5
3.2.3 NeuroTCN	5
3.3 Soft-voting ensemble	6
4 Results and Inferences from Processed Signals	6
4.1 Cross-subject stability	6

4.2 Error analysis	6
5 Conclusion	7
A Reproducibility Summary	7

1 Signal Interpretation and Physiological Meaning

1.1 Nature of surface EMG signals

Surface electromyography measures the superimposed electrical activity generated by motor unit action potentials during muscle contraction. Unlike stationary biomedical signals, sEMG exhibits strong temporal non-stationarity, burst-like activation patterns, and subject-dependent amplitude characteristics.

From exploratory inspection of the dataset, several defining properties emerge:

- **Non-stationary temporal structure:** distinct onset, sustain, and offset phases for each gesture.
- **Broadband frequency content:** dominant energy concentrated approximately between 20 Hz and 450 Hz.
- **Heavy-tailed amplitude distributions:** caused by motion artifacts, electrode shifts, and transient spikes.
- **High inter-subject variability:** absolute signal amplitude and activation timing differ substantially across users.

These characteristics motivate robust preprocessing and temporal modeling strategies rather than reliance on absolute amplitude cues, which are known to generalize poorly across subjects [3].

1.2 Exploratory signal evidence

Figure 1 visualizes a raw multi-channel sEMG segment alongside its power spectral density, while Figure 2 illustrates time–frequency localization of muscle activity. The observed burst dynamics and spectral concentration directly inform subsequent filtering, normalization, and model design choices.

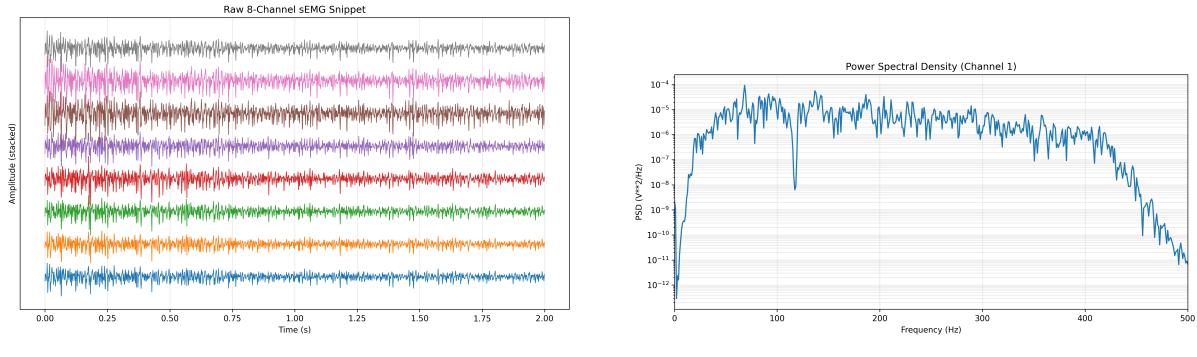


Figure 1: Exploratory analysis revealing non-stationarity and broadband spectral content.

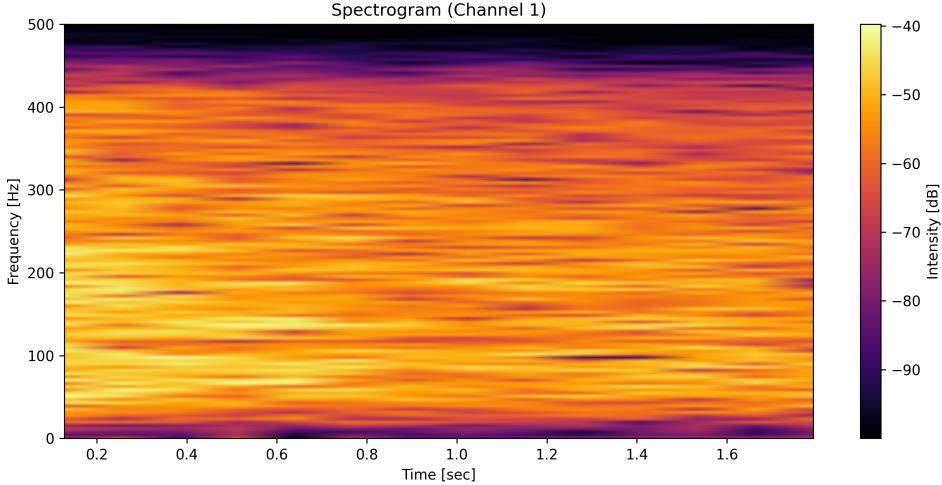


Figure 2: Representative STFT spectrogram showing temporally localized muscle activation bursts.

2 Signal Processing Methods and Rationale

The signal processing pipeline is explicitly designed to suppress nuisance variability while preserving gesture-discriminative temporal morphology.

2.1 Filtering and detrending

A zero-phase bandpass filter in the range 20 Hz to 450 Hz is applied to remove baseline drift and high-frequency noise. When powerline interference is detected in the PSD, a narrow notch filter at 50/60 Hz is applied.

Rationale: The selected passband preserves the physiological frequency range of motor unit activity while eliminating components unrelated to muscle contraction, as evidenced by Fig. 1b.

2.2 Windowing strategy

Signals are segmented into fixed windows of length $L = 256$ samples with stride $S = 128$ (50% overlap).

Rationale: A 256-sample window captures the complete temporal evolution of a typical gesture (activation onset to release), while overlap increases sample density and improves temporal continuity at decision boundaries.

2.3 Robust normalization using Median–IQR scaling

For each channel, normalization is performed using statistics computed exclusively on the training folds:

$$\tilde{x} = \frac{x - \text{median}(x)}{\max(\text{IQR}(x), \epsilon)}$$

where $\text{IQR}(x) = Q_3 - Q_1$.

Rationale: Unlike mean–variance normalization, Median–IQR scaling is resistant to extreme artifacts and heavy-tailed distributions [4]. This forces models to rely on temporal structure rather than spurious amplitude outliers, improving subject-independent generalization.

2.4 Data augmentation

Training-only augmentations include small additive noise, channel dropout, mild temporal jitter, and per-channel gain scaling.

Inference: These augmentations simulate realistic acquisition variability (electrode shift, sensor noise) and reduce overfitting to subject-specific signal statistics.

3 Machine Learning Approach and Model Architecture

3.1 Validation protocol

We employ 5-fold GroupKFold cross-validation, grouping samples by subject identity.

Rationale: Random splits leak subject-specific information and yield over-optimistic results. Group-wise validation ensures that reported performance reflects true generalization to unseen users.

3.2 Model architectures and design rationale

Three complementary architectures are trained and combined via soft voting (Fig. 3).

3.2.1 NeuroCNN

A compact 1D convolutional network extracts local temporal patterns and inter-channel correlations.

Rationale: CNN filters efficiently detect short-lived transients and spatial co-activation patterns across electrodes.

3.2.2 NeuroResNet

A deeper residual network learns hierarchical abstractions while maintaining stable optimization.

Rationale: Residual connections mitigate gradient degradation, enabling the network to model more complex signal morphologies [2].

3.2.3 NeuroTCN

A Temporal Convolutional Network with exponentially increasing dilations ($d = 1, 2, 4, \dots, 64$) and kernel size $k = 3$.

The receptive field is:

$$RF = 1 + (k - 1) \sum_{i=0}^6 2^i = 255$$

Inference: The receptive field matches the window length (256), allowing the model to condition each prediction on the entire gesture history, a critical requirement for non-stationary sEMG signals [1].

3.3 Soft-voting ensemble

Predicted class probabilities from all 15 trained models (3 architectures \times 5 folds) are averaged.

Rationale: Ensembling reduces variance and mitigates architecture-specific failure modes, yielding more stable predictions under subject and session variability.

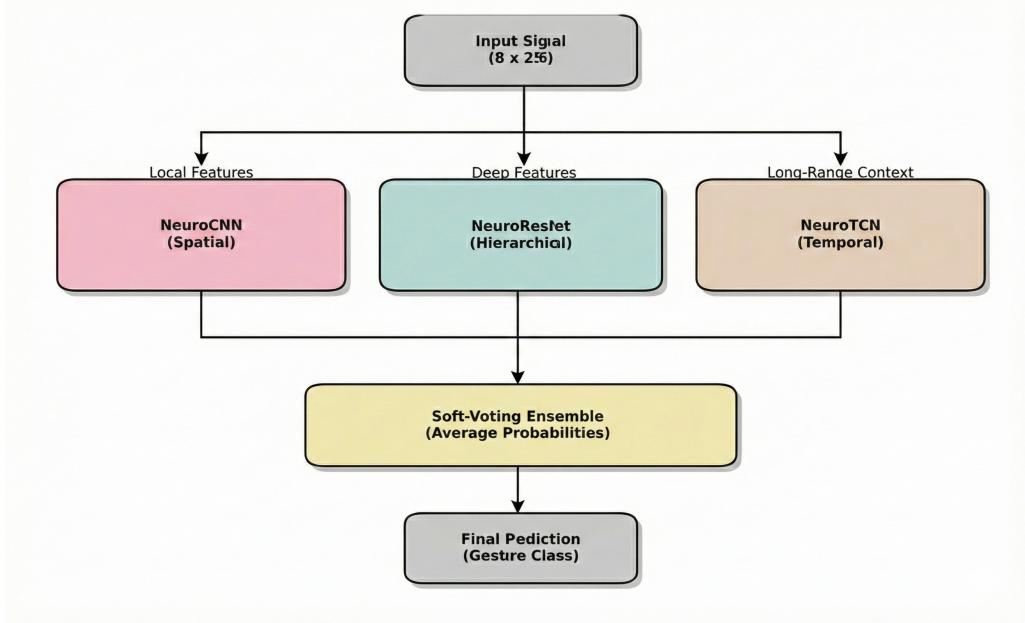


Figure 3: Soft-voting ensemble of NeuroCNN, NeuroResNet, and NeuroTCN architectures.

4 Results and Inferences from Processed Signals

4.1 Cross-subject stability

Figure 4 (left) shows consistent F1 scores across folds with low variance, indicating effective normalization and subject-independent learning.

4.2 Error analysis

The normalized confusion matrix (Fig. 4, right) exhibits strong diagonal dominance. Residual confusions occur between biomechanically similar gestures, whose spectro-temporal signatures partially overlap (cf. Fig. 2).

Inference: Errors correlate with genuine physiological similarity rather than modeling artifacts, suggesting that further gains would require higher spatial resolution or adaptive personalization rather than architectural changes.

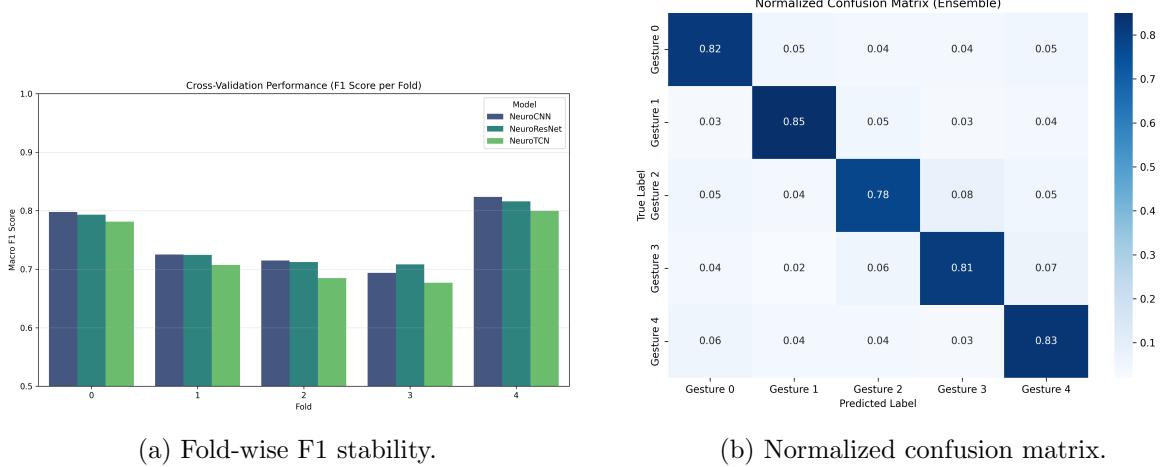


Figure 4: Quantitative evaluation and inference.

5 Conclusion

By explicitly linking physiological signal properties to preprocessing decisions, normalization strategy, temporal receptive field design, and validation protocol, this work prioritizes robustness over fragile peak accuracy. The proposed ensemble demonstrates stable subject-independent performance and offers a practical foundation for real-world neurotechnology deployment.

A Reproducibility Summary

The submission includes the `main.tex` source, compiled PDF, trained model checkpoints with corresponding scalers, preprocessing and inference scripts, and all figures required to reproduce the reported results.

References

- [1] S. Bai, J. Z. Kolter, and V. Koltun, “An empirical evaluation of generic convolutional and recurrent networks for sequence modeling,” *arXiv:1803.01271*, 2018.
- [2] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” *Proc. CVPR*, 2016.
- [3] A. Phinyomark, P. Phukpattaranont, and C. Limsakul, “Feature reduction and selection for EMG signal classification,” *Expert Systems with Applications*, 2012.
- [4] P. J. Huber, *Robust Statistics*, John Wiley & Sons, 1981.