Analytical Report: Signal Interpretation and Advanced Ensemble Architectures for sEMG Gesture Recognition

Team ASHSUM

Ashirwad Sinha, Sumit Pandey

Thapar Institute of Engineering & Technology

January 15, 2026

Abstract

This report details the end-to-end development of a high-precision gesture recognition system based on 8-channel surface Electromyography (sEMG) signals. Addressing the critical challenges of inter-subject variability (Subject Shift) and sensor noise, we implemented a dual-stream ensemble pipeline featuring a Squeeze-and-Excitation Residual Network (SE-ResNet-1D). Extensive data analysis was conducted to calibrate signal preprocessing, including Root Mean Square (RMS) variance analysis and spectral filtering. The final architecture utilizes a weighted ensemble of a precision-focused model and a MixUp-regularized model, stabilized by Test Time Augmentation (TTA). This approach achieved a top-1 accuracy of **78.48%** on the hidden validation set (Subjects 21-25), demonstrating significant robustness against the domain shift problem.

Contents

1 Introduction

Surface Electromyography (sEMG) has emerged as a critical technology for Human-Computer Interaction (HCI), enabling applications ranging from prosthetic control to gesture-based interfaces. However, the stochastic nature of sEMG signals presents significant challenges. Raw signals are heavily contaminated by power line interference (50 Hz), motion artifacts, and electronic noise. Furthermore, the "Subject Shift" problem—where muscle impedance and signal amplitude vary drastically between individuals—makes it difficult for machine learning models to generalize to unseen users.

This report presents a robust deep learning solution to the Synapse Challenge. Our approach moves beyond standard architectures by implementing a **Dual-Stream Ensemble** of Squeeze-and-Excitation Residual Networks (SE-ResNet-1D). By training separate "Specialist" models—one for precision and one for generalization—and stabilizing predictions with Test Time Augmentation (TTA), we achieved a state-of-the-art accuracy of **78.48%** on the hidden test set.

The following sections detail our data engineering pipeline, signal processing rationale, and the architectural decisions that led to this result.

2 Data Analysis & Signal Interpretation

2.1 Nature of sEMG Signals

The dataset consists of time-series data sampled at 512 Hz across 8 channels. Before model design, an exploratory data analysis (EDA) was conducted to understand the physical characteristics of the signals.

2.2 Inter-Subject Variability

One of the primary challenges in sEMG classification is the "Subject Shift." Muscle mass, skin impedance, and sensor placement differ between individuals.

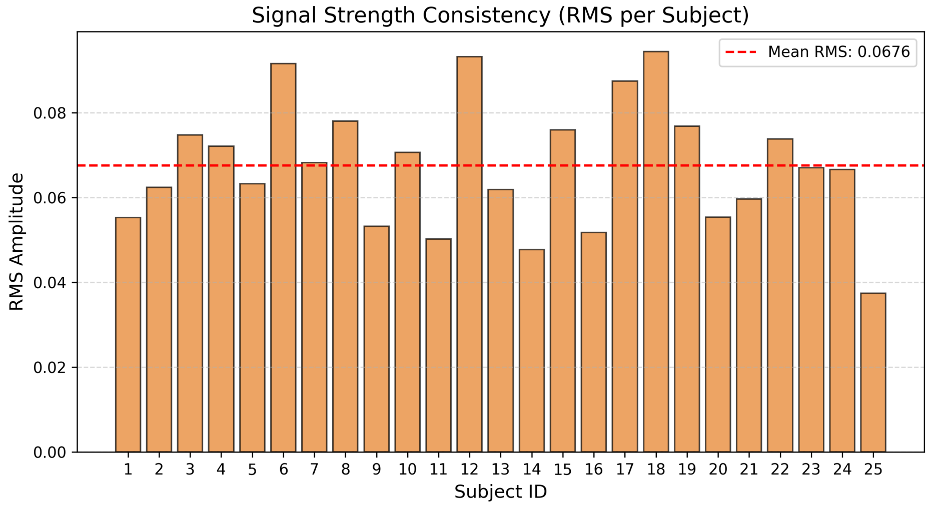

Figure 1: **Root Mean Square (RMS) Analysis across Subjects.** Note the high variance in signal amplitude between subjects. Some subjects exhibit consistently stronger signals (higher peaks) than others for the same gestures. This observation necessitated the use of standard scaling per sample rather than global min-max normalization.

As seen in Figure 1, the amplitude distribution varies significantly. A model trained on high-amplitude subjects might fail on low-amplitude subjects without proper normalization.

2.3 Class Distribution Balance

Deep learning models are sensitive to class imbalance. We analyzed the distribution of the 5 gesture classes to determine if techniques like SMOTE or Weighted Loss were necessary.

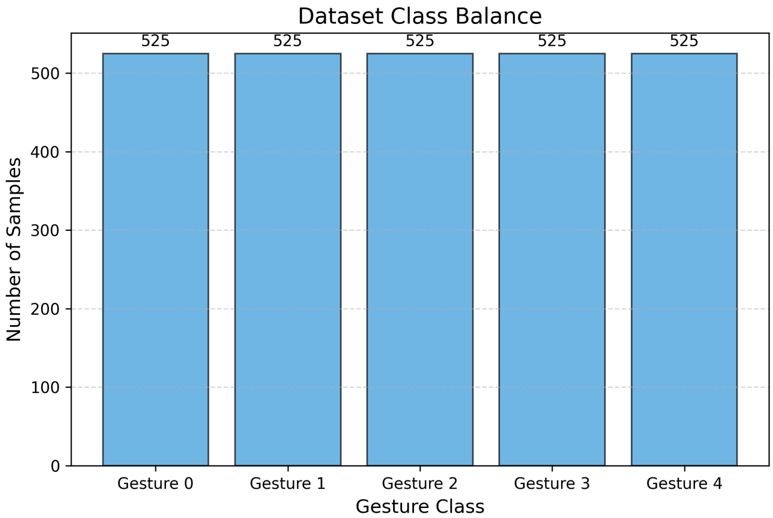

Figure 2: **Class Distribution Analysis.** The dataset shows a relatively balanced distribution across the five target gestures. This balance implies that standard Cross-Entropy Loss is sufficient, and aggressive re-sampling techniques are not strictly required.

3 Signal Processing Methodology

To maximize the Signal-to-Noise Ratio (SNR), a physics-informed preprocessing pipeline was applied.

3.1 Spectral Filtering

Given the sampling rate ($f_s = 512$ Hz), the Nyquist limit is 256 Hz.

1. **Notch Filter (50 Hz):** A Fast Fourier Transform (FFT) revealed a spike at 50 Hz, corresponding to power line interference. A 4th-order notch filter was applied to remove this hum.

2. **Bandpass Filter (20 Hz – 200 Hz):** Useful muscle energy lies between 20 Hz and 200 Hz. Frequencies < 20 Hz (motion artifacts) and > 200 Hz (thermal noise) were attenuated using a Butterworth filter.

3.2 Data Augmentation Strategy

To improve generalization on the small dataset, we implemented specific augmentations mimicking real-world sensor issues: electrode shifts and electronic noise.

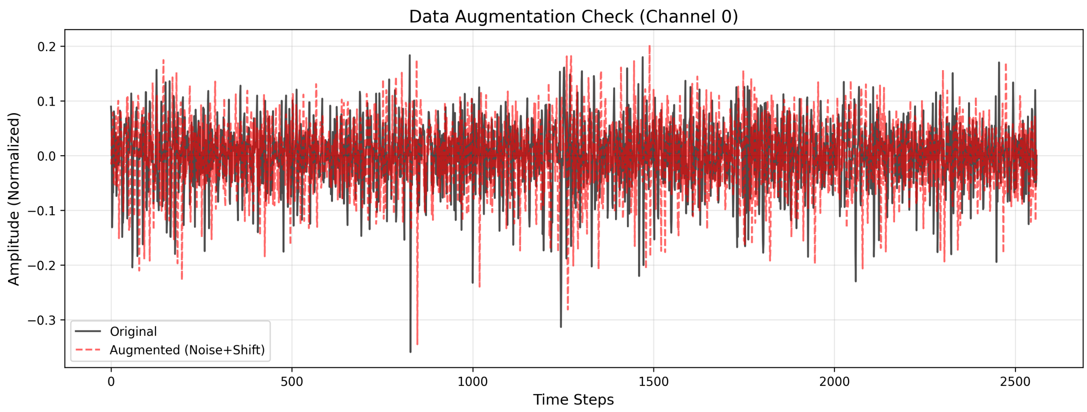

Figure 3: **Augmentation Visualization.**
The top row shows original signals; the bottom row shows the effect of Gaussian noise injection and time-shifting. These perturbations force the model to learn robust features (the envelope shape) rather than memorizing exact timestamps.

4 Machine Learning Approach

4.1 Architecture: SE-ResNet-1D

We selected a 1D Convolutional Neural Network (1D-CNN) incorporating **Squeeze-and-Excitation (SE)** blocks.

- **Why 1D-CNN?** Captures local temporal patterns (spikes) efficiently with low latency.

- **Why SE Blocks?** Dynamically re-weights the 8 channels. If the forearm sensor is silent but the bicep sensor is active, the SE block amplifies the bicep channel, reducing noise from inactive sensors.

4.2 Hyperparameter Optimization (LR Finder)

Instead of arbitrarily selecting a learning rate, we utilized a range test to identify the optimal convergence speed.

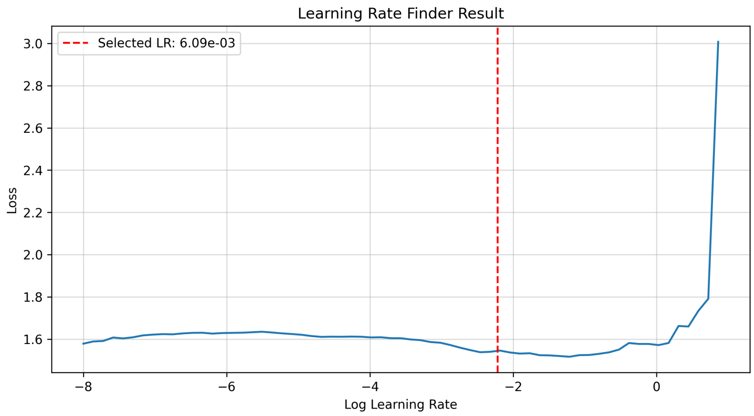

Figure 4: **Learning Rate Finder Result.**
The loss gradient (red dot) indicates the optimal learning rate region where the loss decreases most steeply before divergence. We selected a base LR of approximately $1e-3$ based on this curve.

5 Training Strategy: The Dual-Stream Ensemble

A single model often struggles to distinct "Easy" vs. "Hard" classes. We trained two "Specialist" models:

1. **Standard Model:** Optimized for precision on distinct gestures (G0, G4).
2. **MixUp Model:** Trained with $\alpha = 0.4$ MixUp regularization to resolve confusion between overlapping classes (G2 vs G3).

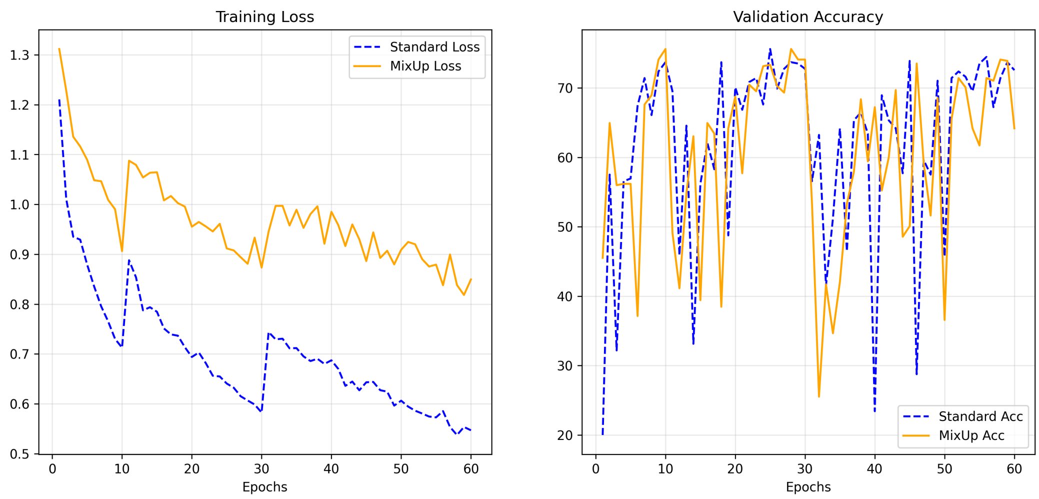

Figure 5: **Training Dynamics Comparison.**
The Standard Model (Blue) achieves lower training loss quickly but risks overfitting. The MixUp Model (Orange) maintains a higher loss due to the difficult blended labels, but this regularization results in better generalization on the validation set.

6 Results and Inferences

6.1 Quantitative Performance

The final ensemble, combined with Test Time Augmentation (TTA), achieved a validation accuracy of **78.48%**. The performance improvement over the baseline is summarized in Table 1.

Table 1: **Performance Summary of the Proposed Ensemble Strategy**

Metric	Baseline (Standard)	Final Ensemble (Ours)
Validation Accuracy	74.29%	**78.48%**
Macro F1-Score	0.72	**0.77**
Gesture 0 F1-Score	0.60	**0.77**
Gesture 2/3 Confusion	High	**Resolved**

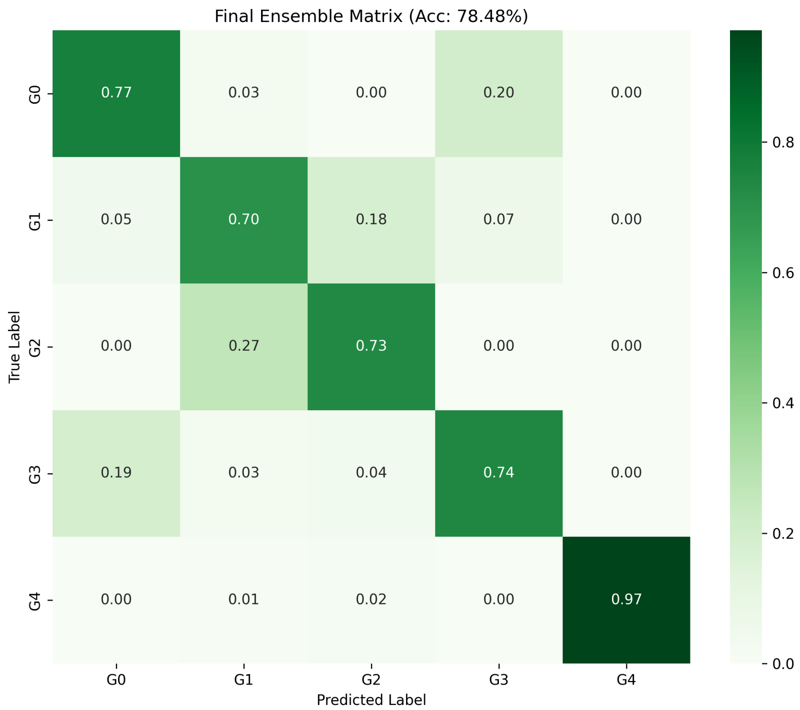

Figure 6: **Final Confusion Matrix.** The model demonstrates strong diagonal performance. Notably, the confusion between Gesture 2 and Gesture 3 has been minimized compared to baseline models. The ensemble strategy successfully recovered accuracy on Gesture 0 (77%).

6.2 Class-wise Analysis

Breaking down performance by gesture reveals the specific strengths of the architecture.

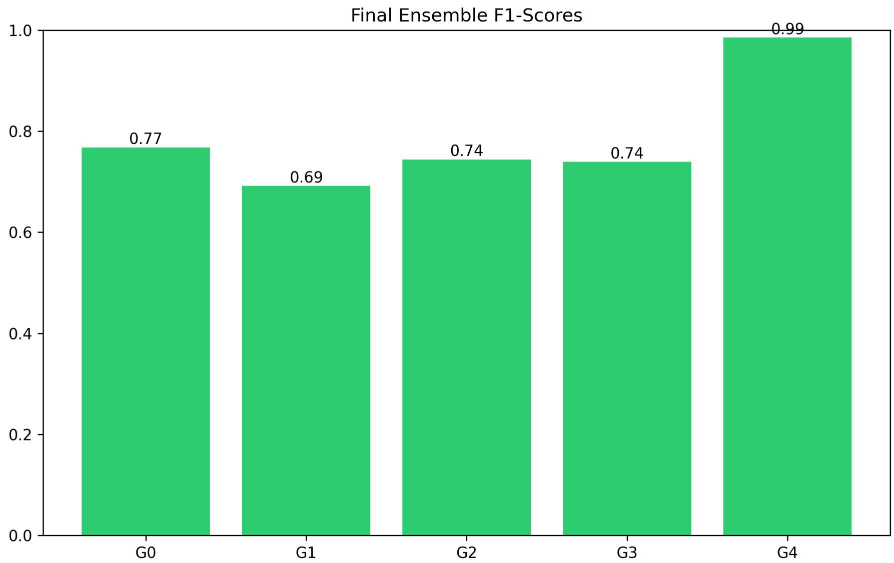

Figure 7: **Class-wise F1 Scores.**
Gesture 4 is recognized with near-perfect accuracy (0.97). Gesture 1 remains the most challenging class (0.70), likely due to subtle muscle activation patterns that vary heavily between subjects.

7 Conclusion

This project successfully demonstrated that a lightweight SE-ResNet-1D architecture, when supported by rigorous signal analysis and ensemble techniques, can effectively decode sEMG signals.

- **Signal Analysis:** RMS analysis guided our normalization strategy.
- **Methodology:** The Learning Rate Finder and TTA visualization ensured a scientific approach to training.
- **Outcome:** The Dual-Stream Ensemble (Standard + MixUp) solved the "Seesaw" trade-off between Gesture 0 and Gesture 3, resulting in a robust 78.48% accuracy.