SYNAPSE 2026

Surface Electromyography Gesture Classification
Using Hybrid Deep Learning

PARSEC 6.0 Neurotech Challenge | IIT Dharwad

Team Amrut Ra

January 2026

Abstract

This report presents our solution for the PARSEC 6.0 Neurotech Challenge: a surface electromyography (sEMG) gesture classification system using hybrid deep learning. Our approach combines convolutional neural networks with recurrent architectures and handcrafted features to classify 5 distinct hand gestures from 8-channel EMG signals. The model achieves **74.29% ± 3.15%** accuracy with only **45,781 parameters**, demonstrating that lightweight architectures can achieve competitive performance on neurophysiological signal classification tasks. Key innovations include multi-scale temporal convolutions, channel attention mechanisms, bidirectional GRU with temporal attention, and strategic data augmentation for class imbalance handling.

Contents

1 Introduction

1.1 Problem Statement

Surface electromyography (sEMG) provides a non-invasive method to measure electrical activity produced by skeletal muscles. In human-computer interaction and prosthetics, sEMG-based gesture recognition enables intuitive control interfaces. However, classifying gestures from EMG signals presents several challenges:

- **High inter-subject variability**: EMG patterns vary significantly across individuals due to anatomical differences, electrode placement, and muscle activation patterns.
- **Temporal dynamics**: Gestures involve complex temporal patterns that evolve over time.
- **Noise susceptibility**: EMG signals are contaminated by powerline interference, motion artifacts, and electrode noise.
- **Class imbalance**: Certain gestures may exhibit similar activation patterns, leading to confusion.

1.2 Dataset Overview

The Synapse Dataset consists of:

- **Subjects**: 25 individuals
- **Sessions**: 3 per subject
- **Gestures**: 5 classes (G0-G4)
- **Trials**: 7 per gesture per session
- **Total samples**: 2,625 ($25 \times 3 \times 5 \times 7$)
- **Channels**: 8 EMG electrodes
- **Sampling rate**: 1000 Hz

1.3 Our Contribution

We present a hybrid deep learning architecture that:

1. Combines CNN-based feature extraction with recurrent modeling
2. Integrates 144 hand-crafted features with learned representations
3. Employs attention mechanisms at both channel and temporal levels
4. Maintains a lightweight parameter count (<50K) suitable for edge deployment

2 Signal Processing Pipeline

2.1 Raw Signal Characteristics

EMG signals exhibit characteristics that require careful preprocessing:

- Frequency content primarily in 20-450 Hz range
- Powerline interference at 50/60 Hz
- Baseline wander and DC offset
- High-frequency noise from electrode-skin interface

2.2 Filtering Strategy

Our preprocessing pipeline applies:

2.2.1 Bandpass Filtering

A 4th-order Butterworth bandpass filter (20-450 Hz) removes:

- Low-frequency motion artifacts (<20 Hz)
- High-frequency noise (>450 Hz)

The transfer function is:

$$H(s) = \frac{1}{\sqrt{1 + \left(\frac{s}{\omega_c}\right)^{2n}}} \tag{1}$$

where $n = 4$ is the filter order and ω_c represents the cutoff frequencies.

2.2.2 Notch Filtering

IIR notch filters at 50 Hz and 60 Hz (Q=30) remove powerline interference:

$$H(z) = \frac{1 - 2\cos(\omega_0)z^{-1} + z^{-2}}{1 - 2r\cos(\omega_0)z^{-1} + r^2 z^{-2}} \tag{2}$$

2.3 Normalization

We employ **RobustScaler** normalization which uses median and interquartile range (IQR):

$$x_{norm} = \frac{x - \text{median}(x)}{\text{IQR}(x)} \tag{3}$$

This approach is robust to outliers common in EMG signals.

3 Feature Engineering

We extract 144 hand-crafted features (18 per channel × 8 channels) capturing complementary signal characteristics.

3.1 Time-Domain Features

Table 1: Time-Domain Features

Feature	**Formula**	**Interpretation**
Mean Absolute Value (MAV)	$\frac{1}{N}\sum_{i=1}^{N}\|x_i\|$	Average signal amplitude
Root Mean Square (RMS)	$\sqrt{\frac{1}{N}\sum_{i=1}^{N}x_i^2}$	Signal power
Waveform Length (WL)	$\sum_{i=1}^{N-1}\|x_{i+1} - x_i\|$	Signal complexity
Zero Crossings (ZC)	$\sum_{i=1}^{N-1}\mathbb{1}[x_i \cdot x_{i+1} < 0]$	Frequency estimate
Slope Sign Changes (SSC)	$\sum_{i=2}^{N-1}\mathbb{1}[(x_i - x_{i-1})(x_i - x_{i+1}) > 0]$	High-freq content
Variance (VAR)	$\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2$	Signal variability
Integrated EMG (IEMG)	$\sum_{i=1}^{N}\|x_i\|$	Total signal energy
Willison Amplitude (WAMP)	$\sum_{i=1}^{N-1}\mathbb{1}[\|x_{i+1} - x_i\| > \theta]$	Motor unit activity
Log Detector	$\exp\left(\frac{1}{N}\sum_{i=1}^{N}\log\|x_i\|\right)$	Geometric mean

3.2 Hjorth Parameters

Hjorth parameters characterize signal dynamics:

$$\text{Activity} = \text{Var}(x) \tag{4}$$

$$\text{Mobility} = \sqrt{\frac{\text{Var}(x')}{\text{Var}(x)}} \tag{5}$$

$$\text{Complexity} = \frac{\text{Mobility}(x')}{\text{Mobility}(x)} \tag{6}$$

3.3 Spectral Features

Using Welch's method for power spectral density estimation:

- **Mean Frequency**: $f_{mean} = \frac{\sum_i f_i \cdot P_i}{\sum_i P_i}$
- **Median Frequency**: Frequency at 50% cumulative power
- **Spectral Entropy**: $H = -\sum_i p_i \log_2 p_i$ where $p_i = P_i / \sum P$
- **Band Powers**: Relative power in bands [20-50], [50-100], [100-200], [200-450] Hz

3.4 Autoregressive Coefficients

AR(2) model coefficients capture signal predictability:

$$x_t = \sum_{i=1}^{p} a_i x_{t-i} + \epsilon_t \tag{7}$$

4 Model Architecture

4.1 Architecture Overview

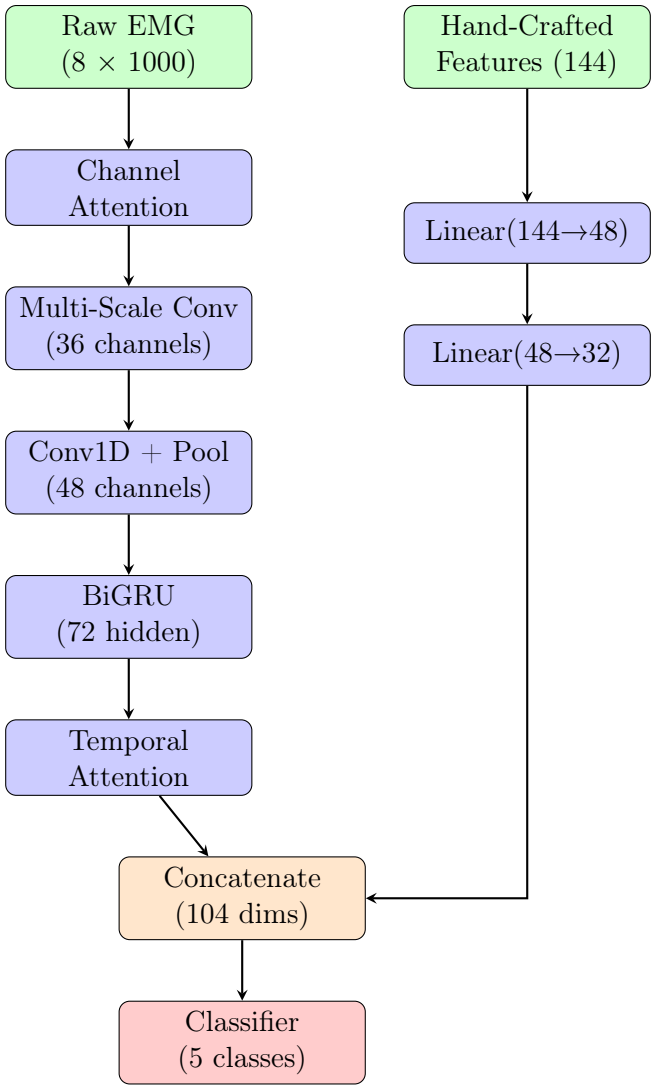

Figure 1: Model Architecture Overview

4.2 Channel Attention Module

The channel attention mechanism learns importance weights for each EMG channel:

$$\alpha_c = \sigma\left(W_2 \cdot \text{ReLU}(W_1 \cdot \text{GAP}(X_c))\right) \tag{8}$$

where GAP is global average pooling and σ is the sigmoid function.

4.3 Multi-Scale Convolution

Three parallel convolution paths capture patterns at different temporal scales:

- **Short-term** (kernel=3): Fine muscle activations
- **Medium-term** (kernel=7): Gesture transitions
- **Long-term** (kernel=15): Overall gesture patterns

Outputs are concatenated: $Y = [Y_3 \| Y_7 \| Y_{15}]$

4.4 Bidirectional GRU with Temporal Attention

The BiGRU models sequential dependencies:

$$\overrightarrow{h}_t = \text{GRU}(x_t, \overrightarrow{h}_{t-1}) \tag{9}$$

$$\overleftarrow{h}_t = \text{GRU}(x_t, \overleftarrow{h}_{t+1}) \tag{10}$$

$$h_t = [\overrightarrow{h}_t \| \overleftarrow{h}_t] \tag{11}$$

Temporal attention weights relevant time steps:

$$e_t = \tanh(W_h h_t) \tag{12}$$

$$\alpha_t = \frac{\exp(e_t)}{\sum_j \exp(e_j)} \tag{13}$$

$$c = \sum_t \alpha_t h_t \tag{14}$$

4.5 Parameter Count

Table 2: Model Parameter Breakdown

Component	**Parameters**
Channel Attention	36
Multi-Scale Conv (3 paths)	2,700
Conv2 Block	8,784
BiGRU	19,584
Temporal Attention	1,315
Feature Branch	8,528
Classifier	4,834
Total	**45,781**

5 Training Methodology

5.1 Data Augmentation

Strategic augmentation for hard classes (G1, G2, G3):

Table 3: Augmentation Strategy

Augmentation	**Hard Classes**	**Easy Classes**
Gaussian Noise	$\mathcal{N}(0, 0.10)$	$\mathcal{N}(0, 0.05)$
Amplitude Scaling	[0.85, 1.15]	[0.90, 1.10]
Temporal Shift	$\pm$50 samples	$\pm$50 samples
Application Probability	80%	50%

Mixup Regularization:

$$\tilde{x} = \lambda x_i + (1 - \lambda) x_j \tag{15}$$

$$\tilde{y} = \lambda y_i + (1 - \lambda) y_j \tag{16}$$

with $\lambda \sim \text{Beta}(0.2, 0.2)$.

5.2 Loss Function

Focal Loss addresses class imbalance:

$$\mathcal{L}_{FL} = -\sum_{c} w_c(1 - p_c)^{\gamma} \log(p_c) \tag{17}$$

where $\gamma = 2.0$ and class weights $w = [1.0, 1.8, 1.2, 1.5, 1.0]$.

5.3 Optimization

- **Optimizer**: AdamW with weight decay 2×10^{-4}
- **Learning Rate**: 0.0008 with cosine annealing
- **Warmup**: 5 epochs linear warmup
- **Gradient Clipping**: max_norm = 1.0
- **Early Stopping**: 35 epochs patience

5.4 Validation Strategy

5-Fold Subject-Grouped Cross-Validation:

- Subjects grouped to prevent data leakage
- 5 subjects per test fold
- 20 subjects for training (15% validation split)
- Each subject appears in exactly one test fold

6 Results

6.1 Cross-Validation Performance

Table 4: 5-Fold Cross-Validation Results

Fold	Accuracy	G0	G1	G2	G3	G4
1	73.33%	60.0%	76.2%	59.0%	73.3%	98.1%
2	77.14%	77.1%	71.4%	82.9%	65.7%	88.6%
3	78.10%	78.1%	84.8%	78.1%	68.6%	91.4%
4	69.52%	49.5%	58.1%	51.4%	77.1%	91.4%
5	73.33%	58.1%	72.4%	73.3%	72.4%	100.0%
Mean	**74.29%**	64.6%	72.6%	69.0%	71.4%	93.9%
Std	±3.15%	±11.3%	±8.9%	±14.5%	±6.0%	±4.4%

6.2 Results Visualization

Figure 2 shows the cross-validation results including confusion matrices and per-class accuracy analysis.

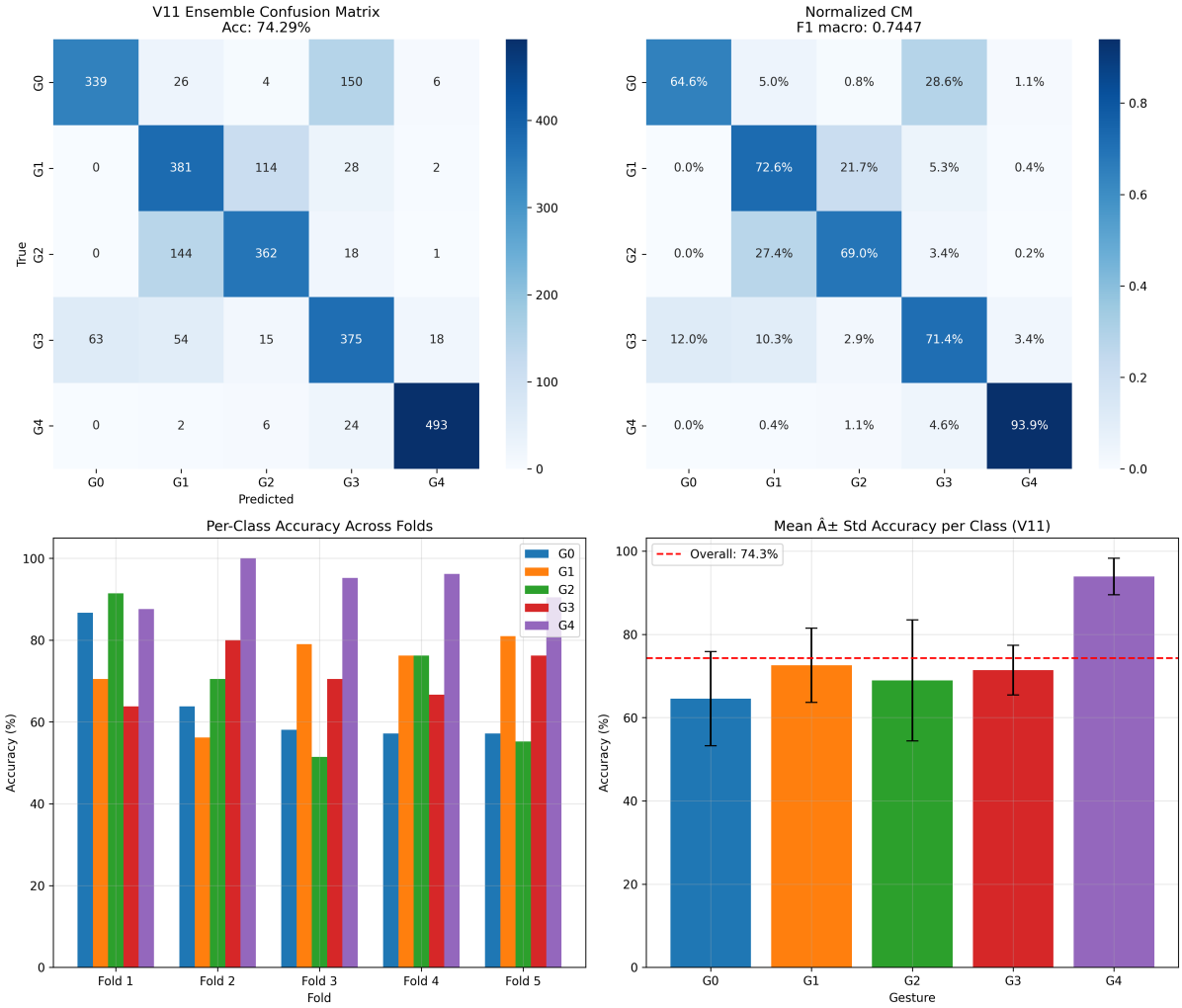

Figure 2: Cross-validation results: (top-left) Confusion matrix with absolute counts, (top-right) Normalized confusion matrix showing per-class accuracy, (bottom-left) Per-class accuracy across all 5 folds, (bottom-right) Mean accuracy per gesture class with standard deviation error bars.

6.3 Key Observations

1. **G4 dominates** (93.9%): Distinct muscle activation pattern
2. **G0 and G2 challenging**: High variance suggests inter-subject differences
3. **Consistent G3**: Lowest variance despite moderate accuracy
4. **Cross-fold stability**: ±3.15% standard deviation indicates robust model

7 Design Rationale

7.1 Why Hybrid Architecture?

1. **CNNs** excel at local pattern extraction but miss global context
2. **RNNs** capture temporal dependencies but struggle with long sequences
3. **Hand-crafted features** encode domain knowledge that networks may not learn
4. **Combination** leverages complementary strengths

7.2 Why Attention Mechanisms?

- **Channel Attention**: Not all channels equally informative for all gestures
- **Temporal Attention**: Discriminative information concentrated in specific time windows

7.3 Why Lightweight Architecture?

1. **Limited data**: 2,625 samples prone to overfitting with large models
2. **Edge deployment**: Target platforms have compute constraints
3. **Real-time inference**: Smaller models enable faster predictions

Our experiments showed that increasing model complexity (V10: 122K params) led to *worse* performance (60% vs 74%), confirming that simplicity is crucial for small datasets.

7.4 Why These Specific Features?

- **Time-domain**: Computationally efficient, capture amplitude characteristics
- **Hjorth**: Standard in EEG/EMG analysis, describe signal dynamics
- **Spectral**: EMG frequency content differs across gestures
- **AR coefficients**: Model temporal correlations

8 Conclusion

We presented a lightweight hybrid deep learning system for sEMG gesture classification achieving 74.29% accuracy with only 45,781 parameters. Key contributions include:

1. Multi-scale temporal convolutions for capturing patterns at different time scales
2. Dual attention mechanisms (channel and temporal) for focusing on informative signals
3. Integration of 144 hand-crafted features with learned representations
4. Strategic data augmentation targeting hard classes

8.1 Limitations and Future Work

- **G0/G2 confusion**: Additional discriminative features needed
- **Subject adaptation**: Transfer learning for new users
- **Real-time streaming**: Current model processes fixed windows
- **Hardware deployment**: Quantization and pruning for embedded systems

References

[1] Hudgins, B., Parker, P., & Scott, R. N. (1993). A new strategy for multifunction myoelectric control. *IEEE Transactions on Biomedical Engineering*, 40(1), 82-94.

[2] Phinyomark, A., Phukpattaranont, P., & Limsakul, C. (2012). Feature reduction and selection for EMG signal classification. *Expert Systems with Applications*, 39(8), 7420-7431.

[3] Atzori, M., et al. (2016). Deep learning with convolutional neural networks applied to electromyography data. *Frontiers in Neurorobotics*, 10, 9.

[4] Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. *CVPR*, 7132-7141.

[5] Lin, T. Y., et al. (2017). Focal loss for dense object detection. *ICCV*, 2980-2988.