

Gesture Recognition from Multi-Session sEMG Signals for Synapse: The Neurotech Challenge

Nikhil Prakash, Karandeep Singh, Charitharth Singh, Ayush Chauhan
School of Computing and Electrical Engineering
Indian Institute of Technology (IIT) Mandi, India

Abstract—Surface electromyography (sEMG) enables non-invasive measurement of muscle activity and plays a crucial role in gesture recognition systems for human–computer interaction and neuroprosthetics. This report presents a complete pipeline for classifying hand gestures using 8-channel sEMG recordings collected across multiple subjects and recording sessions. The proposed approach integrates signal preprocessing, temporal windowing, and deep learning-based time-series modeling. Particular emphasis is placed on leakage-free evaluation through session-wise data splitting and on interpreting learned representations from raw bioelectric signals.

Index Terms—sEMG, Gesture Recognition, Time-Series Classification, Deep Learning, Neurotechnology

I. INTRODUCTION

Surface electromyography (sEMG) is a non-invasive technique used to measure the electrical activity generated by skeletal muscles during contraction [1], [2]. By placing electrodes on the surface of the skin, sEMG captures the superposition of motor unit action potentials, providing a direct representation of underlying neuromuscular activity. Owing to its high temporal resolution and ease of acquisition, sEMG has become a fundamental sensing modality in applications such as prosthetic control, rehabilitation, and human–computer interaction [3].

Gesture recognition using sEMG signals aims to infer intentional hand and wrist movements from patterns of muscle activation. Accurate recognition of gestures enables intuitive and natural interaction mechanisms, particularly in assistive technologies where traditional input devices are impractical. However, sEMG-based gesture recognition remains challenging due to signal non-stationarity, inter-subject variability, electrode displacement, and session-dependent changes in muscle activation patterns [2].

The Synapse: The Neurotech Challenge focuses on addressing these challenges by requiring robust classification of hand gestures from multi-session, multi-subject sEMG recordings. Participants are provided with raw 8-channel sEMG signals collected across different days, and models are evaluated on their ability to generalize to unseen recording sessions while strictly avoiding data leakage.

In this work, we propose a leakage-free gesture recognition pipeline that combines signal preprocessing, temporal windowing, and deep learning-based time-series modeling. The approach leverages convolutional, recurrent, and attention-based architectures to capture both local muscle activation patterns and long-range temporal dependencies [4], [5], while enforcing session-wise data splitting to ensure reliable evaluation and generalization across recording sessions.

II. DATASET DESCRIPTION

The dataset used in this work was released for *Synapse: The Neurotech Challenge* at PARSEC 6.0, IIT Dharwad. It consists of multi-session, multi-subject surface electromyography (sEMG) recordings collected for hand gesture recognition. The dataset is designed to evaluate a model’s ability to generalize across recording sessions conducted on different days.

The recordings were collected from **25 subjects**, each performing **five distinct hand gestures**. Data acquisition was carried out across **three separate recording sessions**, where each session corresponds to a different day of data collection. This multi-session setup introduces natural variations due to electrode repositioning, muscle fatigue, and day-to-day physiological changes, making the task more realistic and challenging.

Each gesture was repeated for **seven trials per subject per session**. Every trial consists of **5 seconds** of continuous sEMG data sampled at a rate of **512 Hz**. Signals were recorded using **eight surface EMG electrodes** placed on the forearm, resulting in synchronized 8-channel time-series data with preserved temporal integrity and no dropped samples.

The dataset is organized hierarchically into three top-level directories: `Session1`, `Session2`, and `Session3`. Each session directory contains 25 subject-specific folders following the naming convention `session{id}_subject{id}`. Within each subject folder, individual trials are stored as CSV files named according to the pattern `gesture{id}_trial{id}.csv`. Each CSV file contains raw sEMG samples arranged as rows corresponding to time steps and columns corresponding to the eight electrode channels.

Ground-truth gesture labels are not explicitly stored within the CSV files. Instead, labels are inferred directly from the file naming convention, where the gesture identifier embedded in

⁰Code and trained models are available at: <https://github.com/karancoderg/SynapsusHackathon>

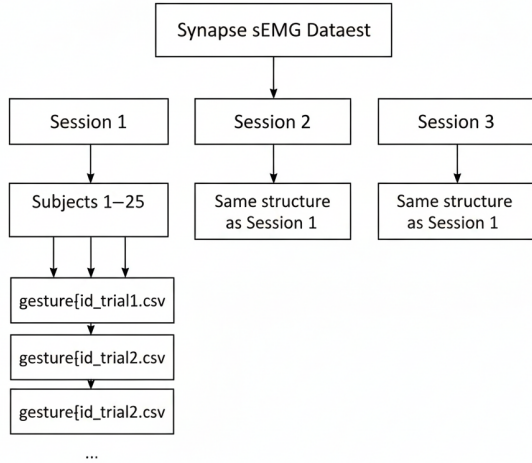


Fig. 1. Hierarchical organization of the Synapse sEMG dataset. The dataset is divided into three recording sessions corresponding to different days. Each session contains recordings from 25 subjects, and each subject folder includes multiple CSV files corresponding to repeated trials of different hand gestures. Sessions 2 and 3 follow the same structure as Session 1.

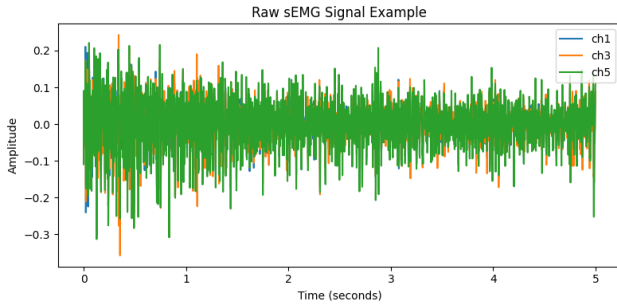


Fig. 2. Example raw sEMG signals from selected channels during a gesture trial, illustrating non-stationary muscle activation patterns.

the filename represents the corresponding gesture class. This design choice ensures that the provided signal data remains raw and unannotated at the sample level, requiring participants to correctly associate labels during preprocessing and model training.

The hierarchical organization of the dataset across sessions, subjects, and gesture trials is illustrated in Fig. 1, highlighting the multi-session and multi-subject nature of the data.

Overall, the dataset provides a comprehensive benchmark for evaluating sEMG-based gesture recognition systems under realistic conditions, with particular emphasis on robustness to inter-session variability and leakage-free evaluation.

III. SIGNAL INTERPRETATION

Surface electromyography (sEMG) signals represent the electrical activity generated by skeletal muscles during volun-

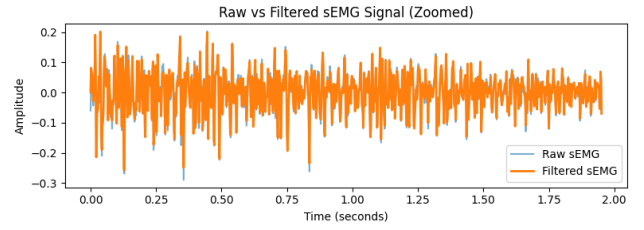


Fig. 3. Illustration of common noise components in sEMG signals, showing raw and band-pass filtered signals. Filtering suppresses low-frequency motion artifacts and out-of-band high-frequency noise while preserving muscle activation patterns.

tary contraction [1]. These signals arise from the superposition of action potentials produced by multiple motor units, where each motor unit consists of a motor neuron and the muscle fibers it innervates. When a gesture is performed, coordinated activation of different muscle groups in the forearm produces characteristic temporal patterns in the recorded sEMG signals across multiple electrodes.

As illustrated in Fig. 2, raw sEMG recordings exhibit rapid fluctuations in amplitude and temporal structure, reflecting the dynamic and non-stationary nature of the underlying muscle activation during gesture execution.

Unlike many conventional biosignals, sEMG signals are inherently *non-stationary*. The statistical properties of the signal, such as amplitude and frequency content, vary over time due to changes in muscle contraction intensity, fatigue, and neuromuscular recruitment strategies [2]. As a result, meaningful gesture-related information is encoded not in individual samples, but in short temporal segments that capture the evolving dynamics of muscle activation.

In addition to temporal variability, sEMG signals exhibit significant *inter-subject variability*. Differences in forearm anatomy, muscle mass, skin impedance, and electrode placement cause the same gesture to produce different signal characteristics across individuals. Furthermore, the dataset incorporates recordings collected across multiple sessions conducted on different days, introducing *inter-session variability* [3]. Day-to-day changes such as electrode repositioning, slight shifts in sensor contact, and physiological conditions further alter the signal distribution. These factors make generalization across sessions a central challenge in sEMG-based gesture recognition.

A. Noise Characteristics

Raw sEMG recordings are affected by several sources of noise and interference that can obscure gesture-related information if left unaddressed. One prominent source is *motion artifacts*, which arise from electrode movement relative to the skin during hand and wrist motion [2]. These artifacts typically manifest as low-frequency fluctuations and can distort the baseline of the signal.

As shown in Fig. 3, motion artifacts primarily affect the low-frequency components of the signal and are effectively attenuated through band-pass filtering. Another common source of

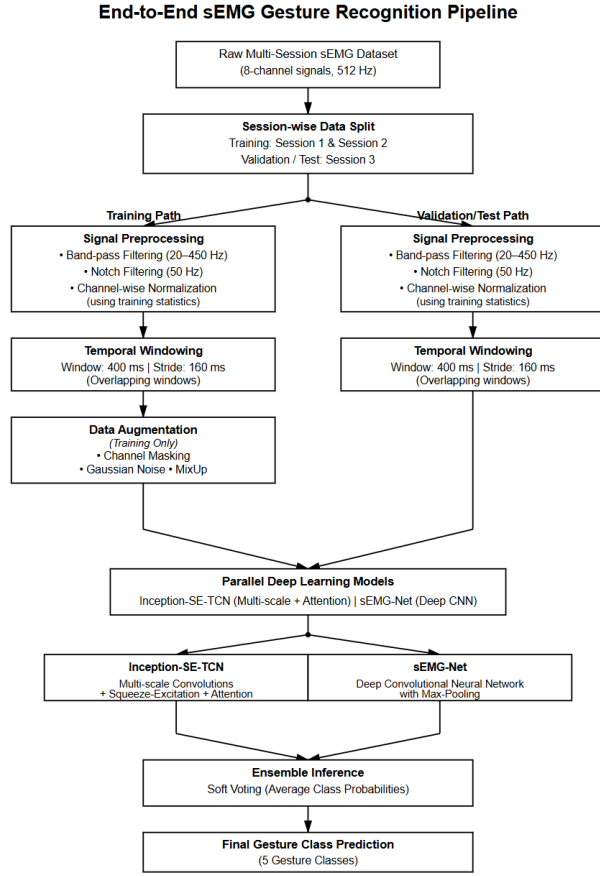


Fig. 4. End-to-end training and inference pipeline for sEMG-based gesture recognition. Raw multi-session sEMG recordings are first split at the session level to prevent data leakage. Signal preprocessing, temporal windowing, and training-only data augmentation are followed by parallel deep learning models whose outputs are combined via ensemble averaging to produce the final gesture prediction.

interference is *power-line noise*, introduced by electromagnetic coupling from surrounding electrical infrastructure. This noise typically appears at the mains frequency and its harmonics, contaminating the frequency spectrum of the sEMG signal [2]. Additionally, *electrode displacement* across sessions leads to changes in the spatial relationship between electrodes and underlying muscles, resulting in variations in signal amplitude and muscle selectivity.

Finally, *crosstalk* occurs when an electrode captures activity from neighboring muscles in addition to the target muscle group [1]. Crosstalk is particularly prevalent in forearm sEMG recordings due to the dense arrangement of muscles, and it contributes to overlapping signal patterns across different gestures. Together, these noise sources motivate the need for robust preprocessing and modeling strategies capable of extracting discriminative temporal patterns while remaining resilient to variability and interference.

IV. SIGNAL PROCESSING PIPELINE

The raw sEMG recordings undergo a structured signal processing pipeline prior to model training, as illustrated in

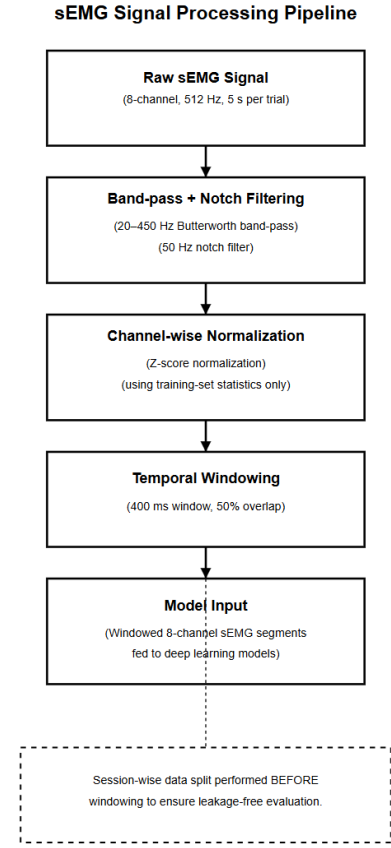


Fig. 5. Overview of the sEMG signal processing pipeline prior to model training. Raw sEMG recordings undergo band-pass and notch filtering, channel-wise normalization using training-set statistics only, and temporal windowing before being provided as model input. Session-wise data splitting is performed before windowing to ensure leakage-free evaluation.

Fig. 4 and Fig. 5. This pipeline is designed to suppress noise, reduce inter-subject variability, and convert continuous recordings into fixed-length representations suitable for machine learning models. The major components of the pipeline include filtering, normalization, temporal windowing, and feature extraction.

A. Filtering

Raw sEMG signals contain frequency components arising from both muscle activity and various sources of noise. Meaningful sEMG information is primarily concentrated in the frequency range of approximately 20–450 Hz, while lower frequencies are often dominated by motion artifacts and baseline drift, and higher frequencies are associated with sensor and electronic noise [2].

To isolate the relevant muscle activation components, a fourth-order band-pass Butterworth filter was applied to each channel, as shown in Fig. 5. In addition, a notch filter centered at the power-line frequency was employed to suppress mains interference [2]. Filtering was performed independently for each channel, ensuring that the temporal structure of the signal was preserved while attenuating unwanted noise components.

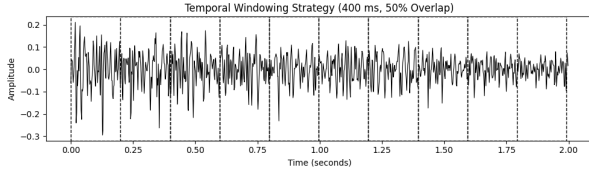


Fig. 6. Temporal windowing strategy illustrating fixed-length windows with overlap.

B. Normalization

The amplitude of sEMG signals varies significantly across subjects and recording sessions due to factors such as muscle mass, skin impedance, electrode placement, and contact quality [3]. Without normalization, these amplitude variations can bias the learning process and hinder generalization.

To address this, channel-wise normalization was applied after filtering, as depicted in Fig. 5. Each channel was standardized using its mean and standard deviation estimated exclusively from the training data. This normalization ensures that all channels contribute comparably during learning and stabilizes optimization, particularly for deep neural network models.

C. Windowing Strategy

sEMG signals are inherently non-stationary, and gesture-related information is distributed over short temporal intervals rather than isolated time samples [2]. Consequently, continuous recordings were segmented into fixed-length temporal windows to capture localized muscle activation patterns, as summarized in Fig. 5 and the overall workflow shown in Fig. 4.

In this work, each trial was segmented using a window length of **400 ms**, corresponding to 205 samples at a sampling rate of 512 Hz. Adjacent windows were generated with a fixed stride of **160 ms**, resulting in overlapping windows that increase the effective number of training samples and enable smoother temporal coverage of muscle activity. Each window thus represents a short sequence of synchronized 8-channel sEMG samples and serves as a single input instance for the models.

As illustrated in Fig. 6, the overlapping windowing scheme ensures that temporal continuity of muscle activation patterns is preserved while enabling dense sampling of gesture dynamics across time.

To prevent data leakage, windowing was performed only after session-level data splitting. All windows extracted from a given trial and session were assigned exclusively to either the training, validation, or test set. This design choice, highlighted explicitly in Fig. 5, Fig. 6, and the end-to-end overview in Fig. 4, ensures that no overlapping or temporally adjacent windows from the same recording appear in multiple splits, enabling a leakage-free evaluation and realistic assessment of cross-session generalization.

Handcrafted Feature Extraction from sEMG Windows

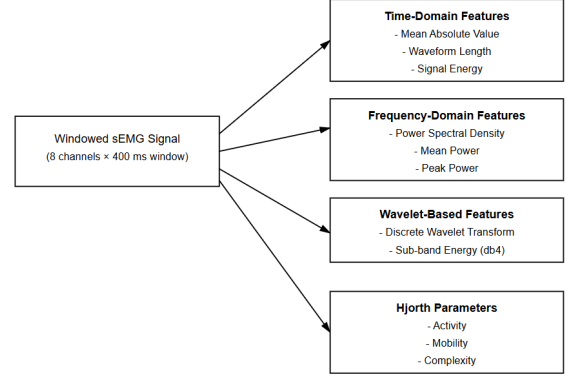


Fig. 7. Categories of handcrafted features extracted from windowed sEMG signals.

D. Feature Extraction

In addition to end-to-end learning from raw windowed signals, handcrafted features were extracted to support classical machine learning models. These features summarize the temporal and spectral characteristics of sEMG signals and provide compact representations that are effective for tree-based classifiers.

As illustrated in Fig. 7, the extracted feature set is organized into complementary categories, including time-domain, frequency-domain, wavelet-based, and Hjorth parameters, each capturing distinct aspects of muscle activation dynamics.

The extracted features include representative time-domain statistics such as mean absolute value, waveform length, and signal energy, as well as frequency-domain descriptors derived from power spectral density estimates [2]. Wavelet-based energy features [8] and Hjorth parameters [7] were also computed to capture multi-resolution and dynamic properties of muscle activation. Such features complement deep learning approaches by explicitly encoding domain knowledge and are particularly effective for models such as gradient-boosted decision trees [9].

V. MACHINE LEARNING METHODOLOGY

This section describes the learning models, optimization objectives, and training strategies employed for sEMG-based gesture classification. The methodology is designed to capture the complex temporal structure of muscle activation signals while ensuring robust generalization across subjects and recording sessions.

A. Model Architecture

Gesture recognition from sEMG signals requires modeling both short-duration muscle activation patterns and longer temporal dependencies spanning an entire gesture execution.

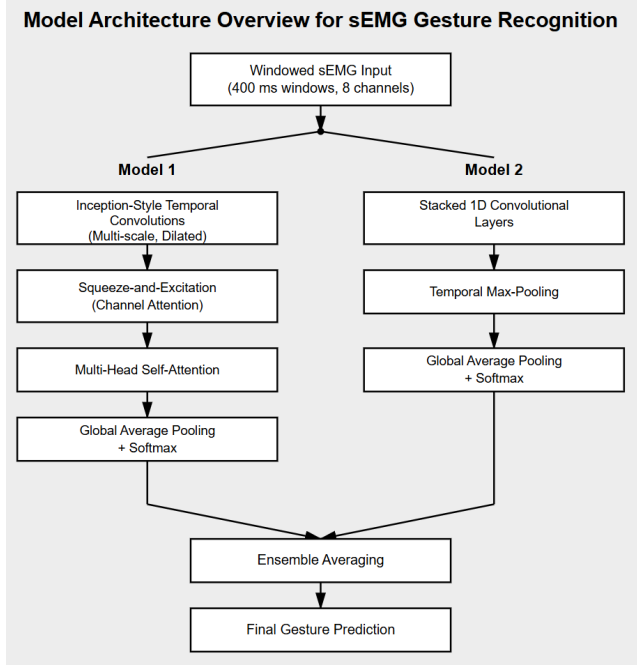


Fig. 8. Model architecture overview for sEMG gesture recognition. Windowed multichannel sEMG inputs are processed in parallel by two complementary deep learning architectures, whose outputs are combined through ensemble averaging to produce the final gesture prediction.

To address this, convolutional and attention-based deep learning architectures were employed, which are well suited for learning hierarchical temporal representations directly from multichannel sEMG windows, as illustrated in Fig. 8.

One-dimensional convolutional neural networks (CNNs) form the backbone of the proposed models, as they efficiently capture local temporal patterns such as rapid changes in muscle activation and channel-wise correlations across short time intervals [4]. Multi-scale convolutional kernels and dilated receptive fields enable the extraction of discriminative features at different temporal resolutions without excessive model depth.

To further enhance temporal modeling, attention-based mechanisms were incorporated within the convolutional frameworks. Multi-head self-attention layers allow the models to dynamically emphasize informative time steps and suppress less relevant signal segments, which is particularly beneficial for non-stationary sEMG data where discriminative information may occur at varying temporal locations [5].

Two complementary architectures were ultimately employed: an Inception-style convolutional network augmented with squeeze-and-excitation and attention mechanisms, and a deep convolutional sEMG-specific network. As depicted in Fig. 8, these architectures operate in parallel on the same windowed input, enabling complementary learning behaviors. Their outputs are later combined through an ensemble strategy to improve robustness and reduce model-specific biases.

Specifically, the Inception-style architecture emphasizes multi-scale temporal feature extraction and adaptive channel

TABLE I
PERFORMANCE COMPARISON OF EXPLORED MODEL ARCHITECTURES DURING MODEL SELECTION.

Model	Accuracy (%)
Transformer	76.70
TCN	78.55
Transformer + TCN (Ensemble)	79.70
XGBoost + CatBoost + LightGBM (Ensemble)	76.50
MAMBA	80.75
GNN-based Model	Overfitting
sEMG-Net	82.70
SE-TCN with Self-Attention	84.10
MAMBA + sEMG-Net + SE-TCN (Ensemble)	84.03
Final Ensemble (SE-TCN + sEMG-Net)	84.55

recalibration, making it effective at capturing fine-grained and transient muscle activation patterns. In contrast, the deeper sEMG-specific convolutional network prioritizes progressive temporal abstraction and global feature aggregation, which is beneficial for modeling sustained activation trends. Combining these architectures through ensemble averaging reduces both variance and inductive bias, yielding more stable and robust predictions than either model alone.

A comparison of the explored model architectures and ensemble configurations during the model selection phase is summarized in Table I. Based on empirical performance and generalization behavior, the SE-TCN with self-attention and the sEMG-specific convolutional network were selected for the final ensemble.

B. Loss Function

Gesture recognition is formulated as a multi-class classification problem, where each input window is assigned to one of the predefined gesture classes. Accordingly, categorical cross-entropy loss with label smoothing was used during training. Label smoothing regularizes the output distribution by preventing overconfident predictions, which improves generalization and stabilizes optimization for deep neural networks.

C. Training Strategy

Model training was conducted with careful consideration of data integrity and generalization, following the strategy illustrated in Fig. 9. To prevent information leakage and ensure realistic evaluation, data were split at the session level prior to windowing, ensuring that recordings from the same day did not appear across training, validation, and test sets. This strategy enforces strict separation between sessions and evaluates the model's ability to generalize across recording days.

Data augmentation was applied exclusively to the training set to improve robustness to signal variability. Augmentation techniques included additive Gaussian noise, channel masking, and MixUp-based interpolation between windowed samples. These operations simulate sensor noise, partial electrode dropout, and inter-sample variability while preserving gesture semantics. No augmentation was applied to validation or test data, as highlighted in Fig. 9.

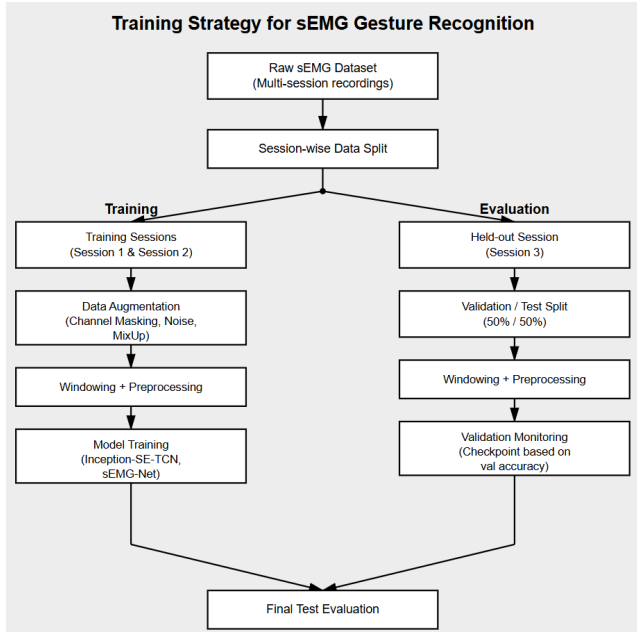


Fig. 9. Training strategy illustrating session-wise data splitting and augmentation applied exclusively to training data. Validation and test data remain untouched to ensure leakage-free evaluation.

Class imbalance and confusion between visually similar gestures were addressed through the use of class-weighted loss optimization, assigning higher weights to underperforming gesture classes during training. This strategy encourages the model to allocate additional capacity to harder classes without altering the data distribution.

Regularization techniques such as dropout, batch normalization, label smoothing, and weight decay were incorporated within the network architectures to reduce overfitting. Training progress was monitored using validation performance, and the best-performing models were selected using checkpoint-based model selection. Together, these strategies promote stable training, prevent overfitting, and enhance generalization across subjects and sessions.

VI. EXPERIMENTAL SETUP AND EVALUATION

This section describes the evaluation protocol, performance metrics, and testing setup used to assess the proposed sEMG-based gesture recognition models. All experiments were conducted under a strict session-wise evaluation framework to ensure leakage-free performance estimation and realistic generalization assessment.

A. Evaluation Protocol

To evaluate cross-session generalization, a session-wise data splitting strategy was employed. Recordings from Session 1 and Session 2 were used exclusively for model training, while all recordings from Session 3 were held out as an unseen test set. No subject overlap or temporal overlap occurs across splits, as windowing is performed only after session-level partitioning.

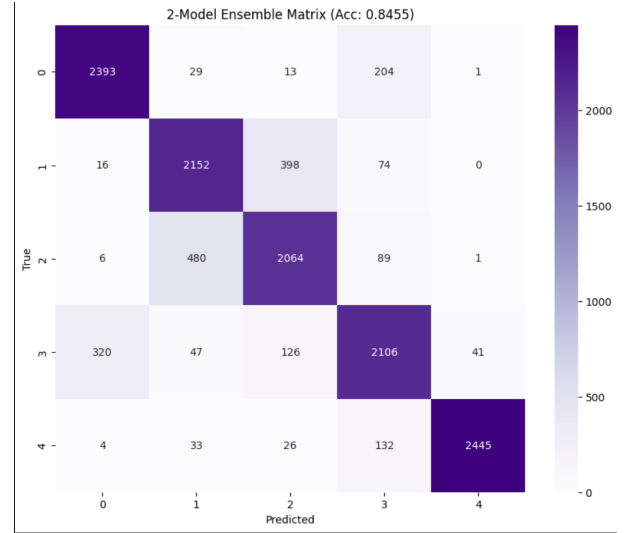


Fig. 10. Confusion matrix for gesture classification on the test set.

Model evaluation was performed at the *window level*. Each test trial was segmented into fixed-length temporal windows of 400 ms with overlap, and each window was independently classified into one of the predefined gesture classes. The final test set consists of **13,200 windowed samples**, each represented as an 8×204 multichannel sEMG segment.

B. Performance Metrics

Model performance was quantified using standard multi-class classification metrics. Overall classification accuracy was computed as the ratio of correctly classified windows to the total number of test windows. In addition, the macro-averaged F1 score was reported to account for class-wise performance balance across gesture categories. Confusion matrices were used to analyze class-specific error patterns and gesture-level confusion behavior.

C. Model Comparison and Ensemble Evaluation

Performance was evaluated for each individual model as well as for the proposed ensemble. Standalone evaluation was performed for the Inception-SE-TCN and sEMG-Net models independently. The ensemble prediction was obtained by simple averaging of the class probability outputs produced by the two models, followed by maximum-probability class selection.

The final ensemble achieves an overall accuracy of **84.84%** and a macro-F1 score of **0.849**, outperforming both individual models. Detailed class-wise performance trends are further analyzed using the confusion matrix shown in Fig. 10.

VII. INSIGHTS AND DISCUSSION

Analysis of the trained models reveals that gesture discrimination is primarily driven by distinctive temporal patterns of muscle activation rather than isolated signal amplitudes. Convolutional layers in the early stages of the network learn localized activation bursts corresponding to short-duration

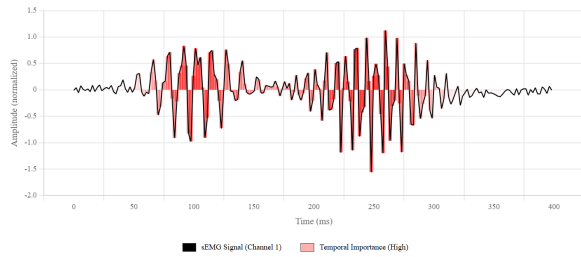


Fig. 11. Visualization of learned temporal importance highlighting discriminative regions of sEMG signals within a 400 ms window.

TABLE II
PERFORMANCE COMPARISON OF INDIVIDUAL MODELS AND ENSEMBLE
ON THE TEST SET.

Model	Accuracy
Inception-SE-TCN	0.8238
sEMG-Net	0.8313
Ensemble (Average)	0.8455

motor unit recruitment, while deeper layers aggregate these responses to form higher-level representations of coordinated muscle activity across channels. Attention-based components further refine this process by assigning greater importance to temporally informative segments within each window, indicating that discriminative gesture cues are not uniformly distributed over time. This behavior is visually supported by the temporal importance patterns illustrated in Fig. 11, where specific time intervals within each window contribute more strongly to the final prediction.

Temporal muscle activation patterns differ in both intensity and coordination across gestures. Gestures involving strong and sustained muscle contractions tend to produce more stable and separable activation profiles, making them easier to classify. In contrast, gestures characterized by subtle or transient muscle movements exhibit overlapping activation patterns across channels, increasing ambiguity and leading to occasional misclassifications. This behavior is clearly reflected in the confusion matrix shown in Fig. 10, where higher confusion is observed between gesture pairs exhibiting similar activation dynamics. These observations highlight the importance of modeling temporal structure and inter-channel relationships in sEMG-based gesture recognition. As summarized in Table II, the ensemble model consistently outperforms both individual architectures, demonstrating the benefit of combining complementary temporal representations.

The multi-session nature of the dataset provides insight into the generalization behavior of the models. While performance remains consistent across sessions, slight degradation is observed when evaluating on recordings from unseen days, reflecting natural inter-session variability. This behavior indicates that the models capture gesture-specific muscle activation patterns while remaining partially sensitive to session-dependent factors such as electrode repositioning and physiological changes. Session-wise training and evaluation play a

critical role in accurately assessing this generalization capability, as evidenced by the structured error patterns observed in Fig. 10.

Despite strong performance, several limitations remain. sEMG-based models are inherently sensitive to electrode placement, and even minor shifts in sensor position can alter the spatial distribution of recorded muscle activity. Additionally, substantial inter-subject variability arising from anatomical and physiological differences limits the extent to which a single model can generalize across users without subject-specific calibration. Finally, real-world deployment introduces further challenges, including sensor noise, inconsistent electrode contact, and variations in movement speed and force, which may not be fully captured in controlled recording conditions. Addressing these limitations is essential for the development of robust and deployable sEMG-based gesture recognition systems.

VIII. CONCLUSION

This work presented a comprehensive pipeline for sEMG-based hand gesture recognition using multi-session, multi-subject recordings. By integrating signal preprocessing, temporal windowing, and deep learning-based time-series modeling, the proposed approach effectively captures discriminative muscle activation patterns while accounting for the non-stationary nature of sEMG signals. A key emphasis was placed on leakage-free evaluation through session-wise data splitting, enabling a realistic assessment of generalization across recording days. The results demonstrate that combining temporal convolutional and attention-based architectures yields robust performance under inter-session variability, highlighting the potential of such methods for practical sEMG-driven human-computer interaction systems.

ACKNOWLEDGMENT

The author thanks the organizers of Synapse: The Neurotech Challenge at PARSEC 6.0, IIT Dharwad, for providing the dataset and evaluation framework.

REFERENCES

- [1] C. J. De Luca, "The use of surface electromyography in biomechanics," *Journal of Applied Biomechanics*, vol. 13, no. 2, pp. 135–163, 1997.
- [2] R. Merletti and P. A. Parker, *Electromyography: Physiology, Engineering, and Non-Invasive Applications*. Hoboken, NJ, USA: Wiley, 2004.
- [3] X. Zhang, X. Zhou, and Y. Chen, "Hand gesture recognition based on surface electromyography," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 1, pp. 23–31, 2013.
- [4] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," *arXiv preprint arXiv:1803.01271*, 2018.
- [5] A. Vaswani *et al.*, "Attention is all you need," in *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, 2017, pp. 5998–6008.
- [6] A. Gulati *et al.*, "Conformer: Convolution-augmented transformer for speech recognition," in *Proc. Interspeech*, 2020, pp. 5036–5040.
- [7] B. Hjorth, "EEG analysis based on time domain properties," *Electroencephalography and Clinical Neurophysiology*, vol. 29, no. 3, pp. 306–310, 1970.
- [8] L. K. Sharma and S. Dandapat, "Wavelet-based feature extraction for EMG signal classification," *Biomedical Signal Processing and Control*, vol. 25, pp. 197–205, 2016.

- [9] G. Ke *et al.*, “LightGBM: A highly efficient gradient boosting decision tree,” in *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, 2017, pp. 3146–3154.

CODE AVAILABILITY

The complete implementation of the proposed preprocessing pipeline, model architectures, and training procedures is publicly available at:

<https://github.com/karancoderg/SynapsusHackathon>