Scholar Publishing Group
International Journal of Big Data Intelligent Technology
https://doi.org/10.38007/IJBDIT.2026.070119
ISSN 2790-0932 Vol. 7, Issue 1: 163-169

Deep Learning for Cross-Subject sEMG Fatigue Classification: Architectures, Window Lengths, and Input Representations

Jun Ye

Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, 15213, United States

Keywords: Surface Electromyography (Semg), Deep Learning, Cross-Subject Classification, Muscle Fatigue Recognition

Abstract. Surface electromyography (sEMG) fatigue estimation has been done through hand-crafted features and traditional machine learning so far. Determine whether end-to-end deep learning directly on sEMG signals has a superior cross-subject baseline and find an optimal model configuration. A well-prepared set of 13 individuals and Leave-One-Subject-Out (LOSO) cross-validation will be used for systematic empirical investigation. In the first round, we will compare XGBoost (a classic machine learning model) with multiple neural network architectures in terms of 2-second and 4-second observation windows: 1D CNN, multi-scale CNN, and LSTM-augmented CNNs. CNN-based models outperform XGBoost in all of the above indices. A single Raw 1D CNN performs better on 2s windows (Macro F1: 0.4798), but a Multi-scale CNN can capture extended patterns in 4s windows more effectively (Macro F1: 0.4786). Adding Recurrent Layers (LSTM) degrades cross-subject generalization. Given the excellent performance of CNN backbones, the second stage will study different input pre-processing techniques, such as rectified signals, envelope integrations and time-frequency representations (STFT, CWT). Counterintuitively, the raw sEMG signals perform better than all explicit preprocessing methods. Rectification and envelope extraction are forms of intervention that reduce the time-domain feature richness and lose some high-frequency MUAP (Motor Unit Action Potential) information. At the same time, time-frequency representations face problems of architectural mismatch and computational bottleneck during full LOSO validation. Based on the above results, we believe that a raw-signal 1D CNN or Multi-scale CNN employing unmodified sEMG windows can be used to perform cross-subject fatigue detection more effectively and simply.

1. Introduction

Fatigue assessment in sEMG helps monitor human performance and rehabilitation for wearable-sensing systems. A typical pipeline in the literature extracts handcrafted time-domain, frequency-

domain or time-frequency features and then uses a classical classifier. Although it is both interpretable and relatively simple, this method explicitly assumes that only certain signal statistics are effective indicators of the start of fatigue.

A deep neural network can be used to learn discriminative temporal structures directly from raw sEMG windows without manual feature engineering. However, continuous fatigue prediction still faces the problem of cross-subject generalisation. Deployable systems are rarely in a position to conduct large-scale labelled-dataset-based training. Therefore, any new architectural or data-processing complexity needs to show real benefits in the strict Leave-One-Subject-Out (LOSO) evaluation.

This paper is about to present an organized, step-by-step experimental study of fatigue detection. We will solve the first three problems.

1. Architectural Paradigm: Does end-to-end raw-window modelling surpass a handcrafted-feature baseline, and are complex sequence models (LSTM) superior to purely convolutional architectures?
2. Temporal Context: How do the performance and rankings of the above architectures change after extending the observation window from 2 seconds to 4 seconds?
3. Given that a strong CNN backbone will be used, what pre-processing strategies should be taken? Time-domain modifications (rectification, envelope) or time-frequency transformations (STFT, CWT) are more fatigue-discriminating than the raw, unmodified signal.

Our research shows that raw-signal learning clearly outperforms handcrafted features in XGBoost. We observe that a Raw 1D CNN peaks at 2s windows, while a multi-scale CNN performs better at 4s windows by using a longer temporal context. An all-encompassing pre-processing experiment shows that the original, unprocessed sEMG waveform is a more stable input. Manual pre-processing operations either reduce the level of detail for time features or add structural complexity that prevents model generalisation among different subjects.

2. Related Work

Previous studies of sEMG analysis have generally taken one of two paths. The first is to use handcrafted features and traditional classifiers, such as Support Vector Machines (SVMs), Random Forests and Gradient Boosting Trees (GBTs) [1]. This Design is still a strong default option because it provides a useful inductive structure and performs well in environments with limited data. As shown in recent systematic reviews [2], the first step of a conventional biosignal pipeline is generally to apply standard signal conditioning and extract robust features, such as RMS, median frequency, or wavelet coefficients.

The second path is a deep neural network that directly learns from sEMG signals and uses 1D CNNs, temporal convolutional models, recurrent modules or hybrid architectures [3]. Although minimalist preprocessing strategies (such as minimally filtered time-series passed to CNNs) have gained attention in gesture recognition [3], their actual effectiveness for continuous fatigue monitoring under various evaluation methods is still unknown. Subject-random splits often overestimate performance, while LOSO evaluation rigorously tests the generalisation capability to new subjects.

Moreover, a relatively large subfield is signal preprocessing. The typical way to obtain the muscle activation envelope in a traditional workflow is full-wave rectification and low-pass filtering [4], and, for instance, Wavelet Transform (WT) or Empirical Mode Decomposition (EMD) can be used to reduce complex artifacts [4]. Although some research has begun converting 1D sEMG

signals into spectrograms or scalograms (CWT) for neural networks [3, 4], they have been applied less frequently in rigorous cross-subject benchmarking for continuous fatigue estimation. Therefore, in this paper, we apply the two preprocessing pipelines to a single CNN backbone for comparison.

3. Methods

3.1 Task Definition

We consider fatigue prediction as an ordinal three-class classification problem with labels 0, 1, and 2. The input of the model is a 4-channel sEMG sliding window. Window-level labels are determined by a majority vote of the aligned 50 Hz fatigue labels in that window.

Evaluation uses Leave-One-Subject-Out (LOSO) cross-validation, and the model is trained on N-1 subjects and tested on the single unseen subject for each fold, then averaged.

3.2 Dataset and Cleaning

A recent study using a dataset with 13 subjects and simultaneous 4-channel upper-limb sEMG recordings and self-reported fatigue [5]. Although this dataset has been mentioned in recent research reviews, to our knowledge, this paper will be the first to build a strong machine-learning benchmark based directly on this dataset for cross-subject fatigue prediction models. According to the exclusion policy proposed by the authors of the dataset [5], the anomalous trials at the full-trial level have been removed. A Cleaning Step will be performed on small-biosignal datasets to address the issue of incorrect or damaged trial data.

3.3 Model Architectures

The four main arrangements are:

Handcrafted Features + XGBoost: The traditional signal-processing baseline using 32-dimensional features (RMS, MAV, mean/median frequency, etc.).

Raw 1D CNN: A three-layer convolutional feature extractor followed by global average pooling and a linear classifier. A basic end-to-end baseline model.

Multi-scale CNN: Parallel convolutional branches of different kernel sizes (e.g., 7, 15, 31) are used to extract multi-resolution temporal features.

Multi-scale CNN + LSTM + Ordinal Head: A recurrent layer is added to learn long-term dependencies, and an ordinal objective function respects the increasing order of fatigue ($0 < 1 < 2$).

3.4 Temporal Context (Window Length)

Two experiments have been carried out using different sliding window sizes to study the impact of a sliding window on the above issue.

2 seconds (75 per cent overlap)

4 seconds (75 per cent overlap)

3.5 Preprocessing Representations

To find the best input representation, we will use many preprocessing methods and test them on the best model. We will use a 1D CNN backbone for the time-domain method. Time-frequency methods are 2D representations, and we map them into a structurally matched 2D CNN to preserve spatial topology. The optimised transforms are:

1. Raw (Baseline): Unmodified raw sEMG signals, normalised to the mean of each trial.

2. Rectified: The absolute value of the signal, xrect[n] = |x[n]|, focusing the network purely on amplitude variations.

3. Raw plus Envelope: Concatenating the raw signal with a low-pass filtered moving-average envelope.

4. Short-Time Fourier Transform (STFT): Provides a uniform time-frequency resolution by sliding a windowed Fourier transform over the signal. The discrete STFT of a signal x[n] is defined as:

STFT(x)[m, k] = ∑∞n=-∞ x[n] w[n - m] e-j(2π/N)kn

where w[n] is the window function (e.g., Hann window), m is the time index, and k is the frequency bin index. We pass the magnitude spectrogram |STFT(x)[m, k]| into the 2D CNN.

5. Continuous Wavelet Transform (CWT): Provides multi-resolution analysis and is relatively high-time-resolution at high frequencies but low-frequency-resolution at low frequencies. The continuous wavelet transform (CWT) of a continuous signal x(t) is given by:

CWT(a, b) = (1 / √|a|) $\int_{-\infty}^{\infty}$ x(t) $\psi((t - b) / a)$ dt∞

where ψ* is the complex conjugate of the mother wavelet (e.g., Morlet), a is the scale parameter (inversely related to frequency), and b is the translation (time) parameter. The resulting scalogram magnitudes are then fed into the 2D CNN.

6. Discrete Wavelet Transform (DWT): Divides a signal into several orthogonal frequency subbands through successive applications of low-pass and high-pass filters. The signal is approximated by the approximation coefficients cAj and detail coefficients cDj at level j. We build the individual components and arrange them spatially for the 2D CNN.

4. Experimental Setup

All the models use per-trial z-score normalisation for training. Optimise the deep learning model with AdamW and use early stopping based on the LOSO validation fold.

The steps of the experimental Design are as follows:

Phase 1 (Architecture and Window): Determine whether raw-window learning outperforms the handcrafted-feature baseline, and find an optimal neural architecture for 2-s and 4-s windows.

Phase 2 (Preprocessing Analysis): Fix the best-performing CNN model and systematically change the input representation to find an optimal data preprocessing method.

Report Macro F1, balanced accuracy, quadratic weighted kappa (QWK), and mean absolute error (MAE). Macro F1 and balanced accuracy assess the performance of class-balanced classification. QWK uses ordinal agreement and MAE represents absolute deviation in the ordinal scale.

5. Results

5.1 Phase 1: 2s Window Benchmark

Table 1. 2s Window Benchmark

Model	Window	Macro F1	Balanced Acc.	QWK	MAE
XGBoost (Handcrafted)	2 s	0.426	0.4634	0.287	0.7057
Multi-scale CNN plus LSTM	2 s	0.3963	0.4565	0.3342	0.6194
Multi-scale CNN	2 s	0.4732	0.5174	0.4562	0.598
Raw 1D CNN	2 s	0.4798	0.5185	0.4389	0.6059

A 1D CNN performs better than XGBoost; thus, both the Macro F1 and QWK increased by 12.6% and 52.9% respectively. The heavily structured LSTM variant performs poorly, and the simplified Multi-scale CNN remains relatively strong.

5.2 Phase 1: 4s Window Benchmark

Table 2. 4s Window Benchmark

Model	Window	Macro F1	Balanced Acc.	QWK	MAE
XGBoost (Handcrafted)	4 s	0.4288	0.4695	0.2859	0.7026
Raw 1D CNN	4 s	0.4545	0.5111	0.4168	0.6303
Multi-scale CNN	4 s	0.4786	0.517	0.4458	0.5885

Extend the window length to 4 seconds and the multi-scale CNN exceeds that of a standard 1D CNN. A multi-branch design is employed to better integrate the extended temporal context without overfitting, and its performance is only slightly lower than that of a standard CNN.

5.3 Phase 2: Preprocessing Representations on the CNN Backbone

A good backbone for CNN has been selected, and a relatively small 2s-window framework was used for preprocessing.

Table 1. Preprocessing Representations on the CNN Backbone

Preprocessing	Backbone	Macro F1	Balanced Acc.	QWK	MAE
Raw (No preprocessing)	1D CNN	0.4798	0.5185	0.4389	0.6059
Rectified	1D CNN	0.4678	0.5052	0.3990	0.6254
Raw plus Envelope	1D CNN	0.3726	0.4217	0.2016	0.7908
CWT	2D CNN	0.4537	0.4884	0.3659	0.6578
STFT	2D CNN	0.4137	0.4440	0.2576	0.7198
DWT	2D CNN	0.3866	0.4283	0.2162	0.7858

The results show that the unmodified raw sEMG signal has the best performance. Every time I have explicitly preprocessed the input, there has been a corresponding decrease in generalisation.

6. Discussion

6.1 CNNs Dominate Classical Methods

End-to-end representation learning from raw windows can also capture physiological fatigue states that traditional statistical features miss. A significant increase in Quadratic Weighted Kappa (QWK) indicates that the CNN can learn the ordinal progression of waveforms autonomously and does not need to be combined with recurrent tracking.

6.2 The Impact of Window Length

The Architecture and the length of the window are not related. A typical 1D CNN works well at 2 seconds. At the 4-second mark, the Receptive Field requirements are altered. The multi-scale CNN has parallel large-kernel convolutions to handle the extended 4-second context and is now the overall top performer. Conversely, adding an LSTM causes catastrophic overfitting in cross-subject evaluation, and it is concluded that a simple, localized sequence learning model is safer than an unbounded recurrent memory for this domain.

6.3 Preprocessing: The Supremacy of Raw Waves

The first important result of our study is that the traditional pre-processing method did not improve the baseline for deep learning.

Time-Domain Feature Dilution: Rectification and envelope concatenation actively degraded the network's capacity for fatigue detection. We assume that removing signal polarity (rectification) or obscuring the high-frequency Motor Unit Action Potential (MUAP) shape by dominant low-frequency amplitude trends (envelope) will lead to a loss of this information. A simple CNN is actually intended to learn these small variations.

In short, the network can learn a good-fitting feature map directly from the original time-series data.

7. Limitations

Models are evaluated in a unified 13-subject cleaned dataset here. LOSO has shown good validation results, and if these results can be extended to external datasets with different sensor hardware and sampling rates, the claims will be strengthened. Future work can also explore specialised lightweight 2D CNN architectures that are designed to reduce the memory footprint of high-dimensional CWT scalograms.

8. Conclusion

This paper systematically presents a benchmark for fatigue-level prediction from sEMG signals and evaluates traditional models, neural network structures, window sizes and input representations. Our results show that a raw-signal CNN is generally superior to handcrafted-feature methods. We find that there are two types of architectures: a relatively simple 1D CNN is suitable for 2-second windows, and a multi-scale CNN can be applied to 4-second windows.

Most significantly, our large-scale preprocessing experiments have strongly recommended avoiding manual signal alteration. Rectification and envelopeing, as well as time-frequency transforms, can reduce the effect of high-frequency physiological indicators or increase the computational load unnecessarily. To achieve good stability, accuracy and generalisation of cross-subject sEMG fatigue detection, the raw and unmodified signal can be fed directly into a well-regularised convolutional network.

References

[1] *Chowdhury, R.H. et al. (2013). "Surface electromyography signal processing and classification techniques." Sensors.*

[2] *Jaramillo-Yánez, A. et al. (2020). "Real-Time Hand Gesture Recognition Using Surface Electromyography and Machine Learning: A Systematic Literature Review." Sensors.*

[3] Li, W. et al. (2021). "Gesture Recognition Using Surface Electromyography and Deep Learning for Prostheses Hand: State-of-the-Art, Challenges, and Future." Frontiers in Neuroscience.

[4] Sul, J.H. et al. (2025). "Electromyography Signal Acquisition, Filtering, and Data Analysis for Exoskeleton Development." Sensors.

[5] Lira et al. (2024). "A Comprehensive Dataset of Surface Electromyography and Self-Perceived Fatigue Levels for Muscle Fatigue Analysis." Sensors (Dataset DOI: 10.3390/s24248081).