PSD-Based Outlier Removal for Enhancing ENG Signal Analysis and Classification

Antonio Coviello
DEIB
Politecnico di Milano
Milan, Italy
antonio.coviello@mail.polimi.it

Silvia Mura
DEIB
Poliecnico di Milano
Milan, Italy
silvia.mura@polimi.it

Andrea Cerutti
DEIB
Politecnico di Milano
Milan, Italy
andrea.cerutti@mail.polimi.it

Umberto Spagnolini
DEIB
Politecnico di Milano
Milan, Italy
umberto.spagnolini@polimi.it

Maurizio Magarini
DEIB
Politecnico di Milano
Milan, Italy
maurizio.magarini@polimi.it

***Abstract*—Accurate classification of electroneurographic (ENG) signals is crucial for applications like prosthetic control and neurorehabilitation. However, noise and artifacts in ENG signals can degrade model performance. To address this, we propose a preprocessing method based on power spectral density (PSD) variance for outlier detection. This approach uses a tunable threshold to balance data retention and removal, effectively eliminating windows with abnormal spectral characteristics. Additionally, we optimize training through early stopping based on the F1-score. This combination of filtering and training optimization enhances ENG signal consistency, improving classification performance by up 8%, particularly for animal experiments with higher levels of noise, while reducing performance variability up to 7 times.**

***Index Terms*—ENG signals, Power Spectral Density (PSD), signal preprocessing, overlapping.**

I. Introduction

Peripheral nerve injuries are a leading cause of chronic disability, significantly affecting patients' quality of life and imposing a burden on healthcare systems [1]. Recent advancements in neural engineering have led to the development of implantable neural interfaces capable of both recording and modulating neural activity. These devices enable the reconstruction of physiological signals and the delivery of targeted electrical stimulation, offering promising therapeutic solutions for restoring lost functions [2]. A notable example is Neural Decode & Stimulate (ND&S) systems, which analyze electroneurographic (ENG) signals to decode motor intentions and dynamically adjust stimulation parameters in real time [3]. ENG signals consist of compound action potentials (CAPs), which are the fundamental units of neural communication, conveying both motor and sensory information, as shown in Fig. 1. CAPs arise from the simultaneous activation of multiple axons within a nerve, and their combined activity forms the ENG signal. However, this collective nature introduces complexity, as overlapping nerve fiber activity generates signals that are highly variable and susceptible to noise [4].

For effective and reliable ENG stimulation, it is crucial to ensure accurate and efficient signal processing and classification. This is especially challenging in real-time applications, such as prosthetic control or closed-loop neurostimulation, where quick responses are essential. The processing pipeline, including signal acquisition, preprocessing, feature extraction, classification, and stimulation command generation, must be completed within roughly 300 milliseconds to preserve natural neuromuscular coordination [5].

To address these challenges, several techniques have been proposed for ENG signal classification [3], [4]. Model-based approaches involve the use of sliding window-based feature extraction, dividing the signal into fixed intervals and calculating statistical descriptors [6]. More advanced machine learning methods, like unsupervised hybrid kernel classifiers, combine clustering and kernel-based learning to extract both local and global features [3]. However, these methods often struggle with noise and artifacts [4].

Deep learning (DL) models, particularly Convolutional Neural Networks (CNNs), have become increasingly popular for ENG signal classification, as they can learn complex patterns from raw data, improving accuracy in real-time applications

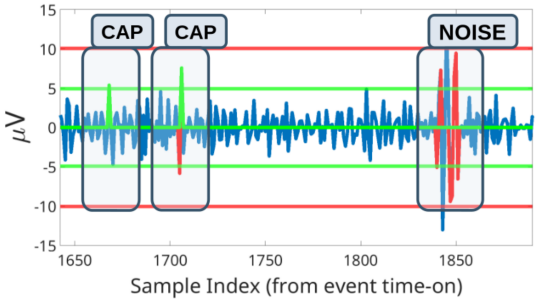


Fig. 1: Representative ENG signal illustrating CAPs embedded with noise.

[3]. However, their sensitivity to noise can result in misclassifications and reduced reliability [6].

Recent research has focused on refining DL methods for ENG analysis. ENGNet, introduced in [3], has shown strong classification performance, with further enhancements, such as data augmentation, improving results and increasing average F1-scores by up to 5% [7]. Additionally, strategies to address class imbalance, such as under- and over-sampling, class weighting, and window overlap, have made models more robust and generalizable [4], [8].

Although these improvements have been made, this study aims to further boost model performance by refining both the training process and the preprocessing strategy. Specifically, in this work we propose:

- An advanced outlier removal strategy, based on the temporal evolution of power spectral density (PSD) variance, is integrated as an additional step within the state-of-the-art preprocessing pipeline in [4]. This approach aims to enhance data quality, improve the overall effectiveness of the pipeline, and mitigate overfitting during model training.
- A training optimization strategy is proposed, employing early stopping based on the F1-score to improve the training process and mitigate the risk of overfitting. For the considered dataset in [9] involving imbalanced class distributions, the F1-score provides a more balanced and informative early-stopping criterion for the training by simultaneously considering precision and recall.

Paper organization: Section 2 details the dataset and the state-of-the-art preprocessing methodology for ENG signals, followed by the spectral analysis and the proposed outlier removal in Section 3. Section 4 describes the ENGNet architecture and the training/testing setup. Section 5 presents the numerical results, and Section 6 concludes the paper.

II. Dataset and State-of-the-art Preprocessing Pipeline

This work analyzes real-time ENG signals recorded from the sciatic nerves of three Sprague Dawley rats during the stimulation of their hind limbs. The dataset, as described in [9], includes four distinct types of sensory activities: dorsiflexion, plantarflexion, touch, and nociception. To ensure experiment reproducibility, stimuli are delivered using an automated system. This system alternates 3 seconds of stimulation with 3 seconds of rest, with reduction to 1 second for pain stimuli. ENG signals are recorded at a sampling frequency of 30 kHz using a multi-contact cuff electrode with $N = 16$ electrodes, arranged across 4 rings with 4 contacts each [9], as shown in Fig. 2. The space-time domain samples of the ENG signal are stored in $\mathbf{X} \in \mathbb{R}^{N\times T}$, with T referring to the number of time samples. The considered dataset in [9] exhibits a class imbalance, resulting in an unequal distribution of class representations.

To process $\mathbf{X}$ efficiently, the pre-processing pipeline detailed in [4] is applied to each signal retrieved from the electrodes in the multi-contact cuff. The state-of-the-art pipeline consists of the following steps :

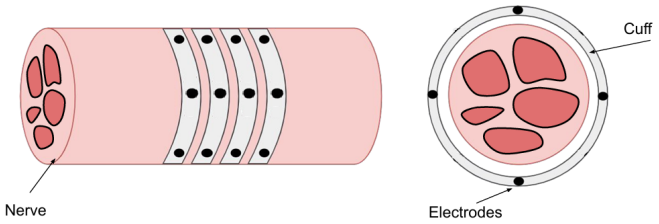


Fig. 2: Multicontact cuff electrode used for ENG signal detection.

- Clamping: to eliminate anomalous voltage peaks, we apply an empirical physiological threshold of 50 μV. Absolute values exceeding this threshold are discarded, resulting in minimal data loss ($< 0.1\%$).
- Bandpass Filtering: An 8th-order Butterworth filter is applied in the 0.8 – 2.5 kHz range to remove low-frequency electromyographic components and reduce high-frequency noise.
- Downsampling: The signals are downsampled to 5 kHz, reducing computational load while retaining key features for classification.
- Signal Segmentation (Windowing): The signal is divided into non-overlapping 100 ms windows. This window length is chosen to contain at least one CAP [10].
- Activity Window Definition: The dataset in [9] provides sensor data that enables the division between activity interval, denoted as $[\mathbf{X}]_{T\text{on}}$, which represent the time periods during stimulus application, and rest interval, denoted as $[\mathbf{X}]_{T\text{off}}$, corresponding to periods without stimulus.

The signal-to-noise ratio (SNR) over all the electrodes is computed as the ratio between the signal power during activity time and that during rest time, defined as

$$\text{SNR} = \frac{1}{\xi}\frac{||[\mathbf{X}]_{T\text{on}}||^2}{||[\mathbf{X}]_{T\text{off}}||^2}, \tag{1}$$

where ξ is a normalization factor accounting for the number of samples in each interval. The average SNR across the stimuli for the considered dataset in [9] is approximately 1.85 dB for Animal 1, 1.30 dB for Animal 2, and 1.15 dB for Animal 3.

Based on these observations, the existing preprocessing pipeline exhibits limitations in handling frequency-domain outliers resulting from noise and artifacts. To overcome this, the following subsection introduces a spectral filtering step as an enhancement to improve the robustness and effectiveness of the pipeline.

III. Spectral Analysis and Filtering of the ENG Signals

The ENG signals contain noisy components that are difficult to filter in the time domain but can be more effectively detected and filtered in the frequency domain. The PSD,

which measures the signal's power distribution across various frequencies, is used to identify dominant components and detect noise or artifacts. Widely applied in EEG and EMG [11], PSD facilitates feature extraction by highlighting key frequency components, differentiates neural activity based on spectral patterns, and identifies artifacts that could disrupt analysis and reduce classification accuracy.

Let the FFT of the ENG signal from the nth electrode in the ith time window with $T_{\text{on,i}}$ samples be denoted as $X_{n,i}(f)$, the PSD of the signal is then defined as:

$$S_{n,i}(f) = \frac{1}{T_{\text{on,i}}} |X_{n,i}(f)|^2. \tag{2}$$

Figure 3 presents the average PSD μ_{PSD} for each of the four experimental conditions, dorsiflexion, plantarflexion, touch, and nociception, for Animal 2. Notably, the power distribution varies across the electrodes of the multi-contact cuff, depending on the applied stimulus. This variation likely reflects the activation of different axonal pathways within the nerve in response to distinct stimuli.

To further evaluate signal quality, the PSD variance across repeated stimuli of the same type for a given animal is computed and shown in Fig.4. Dorsiflexion exhibits anomalies at frequencies below 1000 Hz, whereas plantarflexion shows broader but less intense variations (Fig. 4a), while for the nociceptive condition (Fig. 4d), variance is notably higher between 800 Hz and 1500 Hz, indicating substantial fluctuations in signal content. Touch and plantarflexion conditions generally show lower variance, and localized peaks still appear, particularly in peripheral electrodes.

These observations reveal the presence of noise or artifacts in specific time windows, likely arising from experimental disturbances or intrinsic variability in neural responses. To address this, we introduce a targeted method based on spectral analysis, which selectively removes only the corrupted signal segments, thereby preserving the useful portions of each stimulus. This approach enhances dataset quality by eliminating noisy data, improves model robustness by preventing the learning of artifact-driven patterns, and increases result interpretability by reducing uncertainty in the analysis. The outlier detection and removal procedure is structured as follows:

- PSD Calculation: For each activation window in Sec. II, the PSD is computed as in (2).
- Outlier Detection and Removal The mean $\mu_{\text{PSD},n}(f)$ and standard deviation $\sigma_{\text{PSD},n}(f)$ of the PSD at each frequency f are calculated across all the activity windows. The frequency component f in window i and electrode n is flagged as an outlier if it satisfies the condition:

$$S_{n,i}(f) > \mu_{\text{PSD},n}(f) + K \cdot \sigma_{\text{PSD},n}(f), \tag{3}$$

where K represents a sensitivity threshold: higher values allow for greater variance, while lower values increase the number of filtering points. This detection and removal process is applied over three rounds to ensure convergence and the stabilization of the filtered dataset.

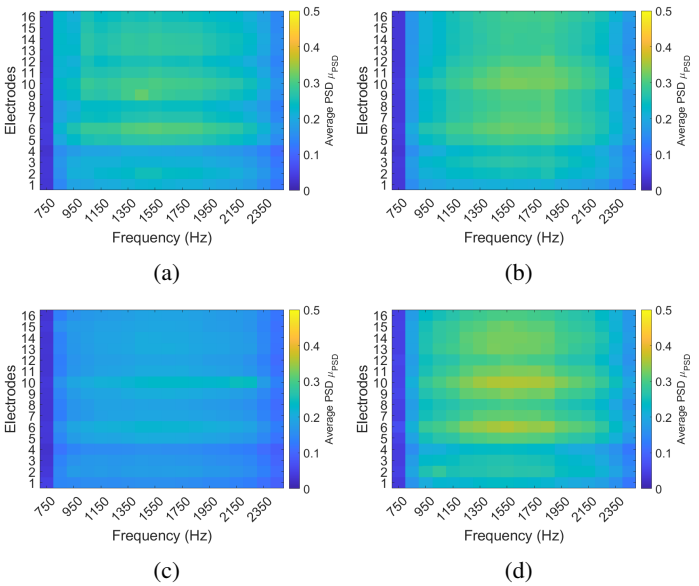


Fig. 3: Average PSD for each of the four experimental conditions: (a) dorsiflexion, (b) plantarflexion, (c) touch, and (d) nociception, for Animal 2.

The impact of outlier removal is evident in Fig. 5, which illustrates the variance of the PSD for dorsiflexion and nociception in Animal 2. A comparison with Fig. 4a and 4d reveals a noticeable reduction in the PSD variance following outlier removal, indicating an increase in the consistency of the signals within the same class.

The proposed outlier removal method is integrated into the state-of-the-art preprocessing pipeline, yielding the filtered

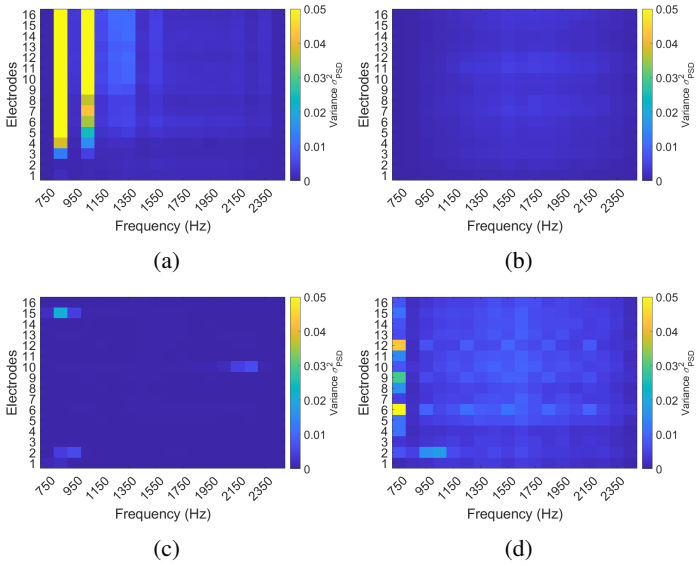


Fig. 4: Variance of the PSD for each of the four experimental conditions: (a) dorsiflexion, (b) plantarflexion, (c) touch, and (d) nociception, for Animal 2.

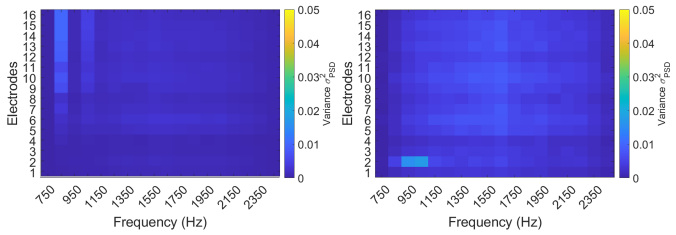


Fig. 5: Variance of the PSD for (a) dorsiflexion and (b) nociception stimuli after removing windows containing outliers, for Animal 2.

ENG signals, denoted as $\tilde{\mathbf{X}}$, which are then used as input to the CNN network for ENG stimuli classification.

IV. ENGNet Training and Test Set Definition

This section outlines the CNN architecture used for ENG signal classification, along with the definition of the training and test datasets.

A. ENGNet

ENGNet is a compact deep learning architecture tailored for the real-time classification of ENG signals [3]. The architecture's input is the matrix $\tilde{\mathbf{X}}_i \in \mathbb{R}^{N \times T_{\text{on},i}}$, which refers to the ith activity window.

The architecture aims at minimizing the training loss function, defined as the categorical cross-entropy loss function

$$L(\tilde{\mathbf{X}}_i) = -\sum_{c \in \mathcal{C}} p(\tilde{\mathbf{X}}_i) \log(\hat{p}_c(\tilde{\mathbf{X}}_i)) \tag{4}$$

where $\mathcal{C}$ represents the set of classes, i.e, dorsiflexion, plantarflexion, touch and nociception, $p(\tilde{\mathbf{X}}_i)$ is the true class label, represented as a one-hot encoded vector (i.e., for the correct class, $p(\tilde{\mathbf{X}}_i) = 1$, and for all other classes, $p(\tilde{\mathbf{X}}_i) = 0$) and $\hat{p}_c(\tilde{\mathbf{X}}_i)$ is the probability assigned by the model that the sample belong to class c.

ENGNet's architecture employs 1D temporal and spatial convolutions to extract frequency-specific and channel-selective features, thereby enabling efficient processing of the ENG data. The network comprises two primary convolutional blocks: the first captures temporal and spatial patterns through depthwise convolutions, while the second utilizes separable convolutions to reduce complexity and effectively combine feature maps. Each block incorporates normalization, exponential linear unit activation, average pooling, and dropout layers to enhance learning and generalization. The final features are passed through a dense layer and a softmax classifier, yielding the predicted class.

To evaluate ENGNet's performance in ENG signal classification, two key metrics are employed: accuracy and F1-score. Accuracy measures the proportion of correct predictions out of the total number of cases, providing an overall assessment of the model's performance. However, in scenarios involving imbalanced datasets, where certain classes are underrepresented, accuracy may not fully capture the model's effectiveness. In such cases, the F1-score becomes particularly relevant, as it balances precision (the proportion of true positive predictions among all positive predictions) and recall (the proportion of true positive predictions among all actual positives) into a single metric. This balance is crucial for identifying potential class imbalances and ensuring that the model performs well across all classes. In the results section, we will compare these metrics across different preprocessing configurations to provide a comprehensive evaluation of ENGNet's classification capabilities.

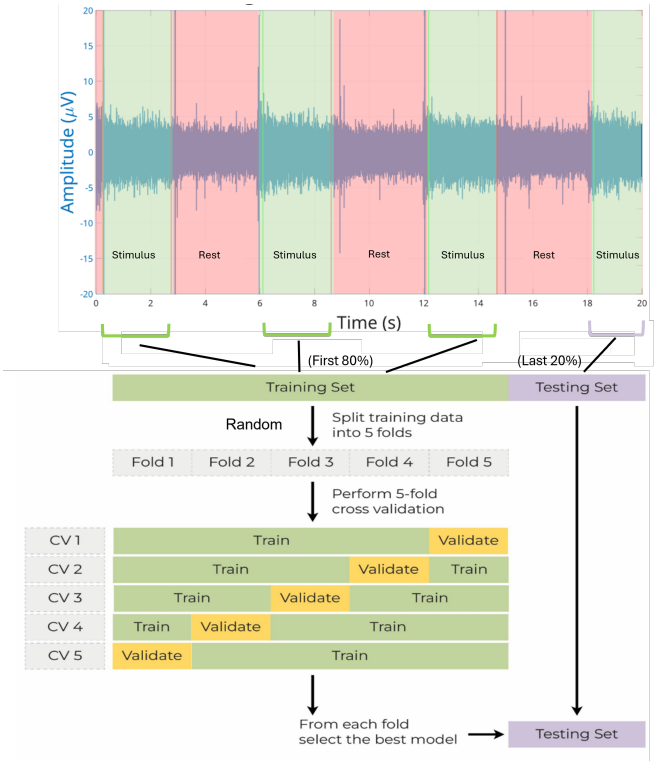


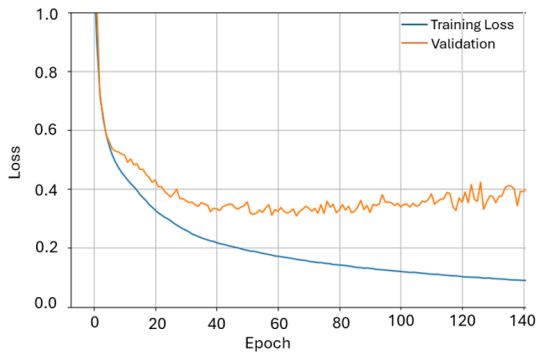
Fig. 6: Dataset division and Training set definition.

B. Training and Test Set Definition

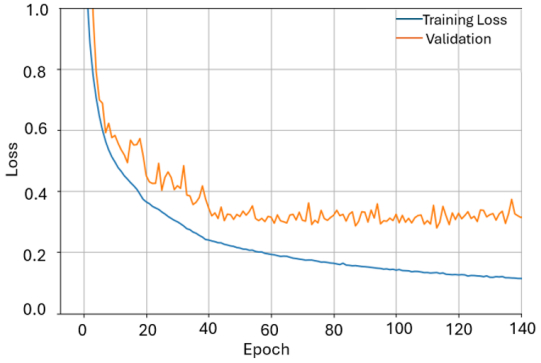
The dataset is segmented into 100 ms activation windows, as described in Sec. II, and divided into training (80%) and test (20%) sets, ensuring complete independence of the test set. The training set consists of the initial signal segments, while the test set includes the later segments, preserving an 80:20 ratio. This setup simulates a dynamic environment where stimuli may evolve over time, as shown in Fig. 6.

The outlier filtering procedure, detailed in Sec. III, is applied only to the training set to prevent bias in performance evaluation. Training is conducted with a random 5-fold split within the training set, maintaining an 80:20 training-validation ratio.

Two stopping criteria are evaluated: a fixed limit of 150 epochs, as used in [3], and the maximization of the F1-score, which is proposed in this work and is commonly preferred for imbalanced datasets due to its ability to balance precision and recall, offering a more reliable measure of performance.



(a)



(b)

Fig. 7: Training loss function for the ENGNet with preprocessing in [4] and with the add of PSD-based outlier removal, for Animal 3.

To assess the network's behavior during training, the training and validation losses are monitored at each epoch, providing insights into potential underfitting or overfitting. Specifically, overfitting is identified when the validation loss increases while the training loss continues to decrease.

Figure 7 illustrates the effect of the PSD-based outlier removal on the training loss of the ENGNet model, as defined in [3]. The outlier removal process mitigates overfitting by eliminating out-of-distribution samples, as shown in Fig. 7b. In contrast, training without outlier removal (Fig. 7a) displays typical overfitting, likely due to the model learning from anomalous samples.

V. Numerical Results

This section evaluates the effect of the parameter K on the percentage of anomalous windows removed from the ENG signals and identifies the optimal K configuration. It also compares the classification performance of the proposed method, which integrates PSD-based outlier removal and F1-score-based training early stopping, with the state-of-the-art approach presented in [3], which employs the preprocessing pipeline from [4] and uses a fixed training duration of 150 epochs.

The outlier analysis in ENG signals enabled the identification and removal of anomalous windows, thereby improving data quality and enhancing subsequent analysis accuracy. The number of outliers removed is highly dependent on the K parameter used during classification. Table I shows the percentage of outlier windows removed per animal and class for different K values. As K decreases, the removal rate increases: for $K = 6$, removal rates remain below 6%; for $K = 5$, the touch class approaches 10%; and for $K = 4$, removal exceeds 11% in some cases.

Animal	Class	K = 4	K = 5	K = 6
Animal 1	Noci	5.26 %	2.89 %	1.61%
	Dorsi	2.83%	1.19 %	0.72%
	Plantar	10.03%	5.95 %	4.67%
	Touch	11.36%	9.59 %	1.88%
Animal 2	Noci	4.14%	1.50 %	1.26%
	Dorsi	6.72%	3.07 %	2.35%
	Plantar	3.36%	0.85 %	0.16%
	Touch	9.01%	6.76 %	6.57%
Animal 3	Noci	2.95%	0.55 %	0.55%
	Dorsi	1.84%	0.53 %	0.35%
	Plantar	4.24%	0.99 %	0.32%
	Touch	3.08%	1.28 %	0.88%

TABLE I: Percentage of eliminated window of 100 ms. The total number of windows per class is as follows: dorsiflexion 3000; plantarflexion 3000; touch 2000; nociception 500.

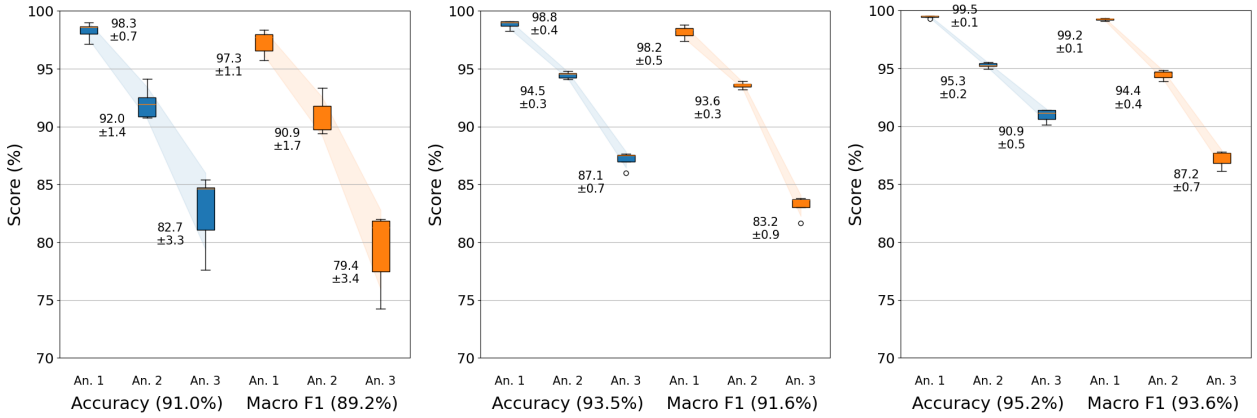
The touch class consistently exhibits the highest removal rates, suggesting greater sensitivity to outlier detection, while the nociception class shows the lowest rates, likely due to its smaller dataset size and lower susceptibility to noise. Furthermore, inter-animal variability is observed, correlating with differences in SNR values, as discussed in Sec. II.

TABLE II: Accuracy and F1 score (%) for different values of K, averaged over the animals.

Setting	Metric	An. 1	An. 2	An. 3
$K = 4$	Accuracy	98.8 ± 0.4	94.2 ± 1.0	86.1 ± 1.1
	F1-score	98.4 ± 0.5	93.1 ± 1.4	81.9 ± 1.6
$K = 5$	Accuracy	98.8 ± 0.4	$\mathbf{94.5 \pm 0.3}$	$\mathbf{87.1 \pm 0.7}$
	F1-score	98.2 ± 0.5	$\mathbf{93.6 \pm 0.3}$	$\mathbf{83.2 \pm 0.9}$
$K = 6$	Accuracy	98.8 ± 0.4	94.2 ± 1.0	86.1 ± 1.1
	F1-score	98.4 ± 0.5	93.1 ± 1.2	81.9 ± 1.6

The optimal value of K is identified by comparing the classification performance of the ENGNet across various configurations. Table II presents the average accuracy and F1-score values, along with their standard deviations, for each setting. Notably, $K = 5$ delivers the best performance, especially for Animal 2 and Animal 3, which exhibit lower SNR values. This configuration optimizes the model's performance for the animals in the dataset, reducing the standard deviation by nearly half compared to other configurations.

Finally, the proposed method, which combines optimized outlier removal with $K = 5$ and early stopping guided by the F1-score, is compared to the baseline approach from [3]. Figure 8 shows the accuracy and F1-score performance for each animal, averaged across classes, for [3] (Fig. 8a), proposed method (Fig. 8b). The proposed method improves



(a) (b) (c)

Fig. 8: Accuracy and F1-score performance per animal and averaged across classification classes: (a) [3], (b) proposed method, (c) proposed method with 80% window overlap.

the benchmark [3] by 2.5% and 4.5% in both accuracy and F1-score, while reducing the standard deviation by approximately one-fourth. The improvement is most noticeable for Animal 2 and Animal 3, which have lower SNR values, whereas Animal 1 shows a slight performance improvement.

The inclusion of window overlap has been confirmed to enhance ENG classification performance, as detailed in [4]. Therefore, Fig. 8c shows the proposed method with 80% window overlap (Fig. 8c), which outperforms [3] by around 2% for Animal 1, 4% for Animal 2, and 8% for Animal 3 in both accuracy and F1-score. Additionally, the standard deviation for Animals 2 and 3 is further reduced by approximately 7 times for accuracy and 5 times for the F1-score.

In conclusion, the proposed optimized method, incorporating outlier removal with $K = 5$ and early stopping based on the F1-score, enhances ENG signal classification. The method improves performance by up to 4.5% without overlap and up to 8% with overlap, particularly for animals with lower SNR. Additionally, it outperforms the state-of-the-art [3], improving classification performance and reducing variability.

VI. Conclusions

Accurate classification of electroneurographic (ENG) signals is vital for prosthetic control and neurorehabilitation, but noise and artifacts can degrade model performance. To address this, we propose a preprocessing method using power spectral density (PSD) variance for outlier detection, coupled with training optimization via early stopping based on the F1-score. This combination enhances signal consistency, leading to performance improvements of up to 8%, especially for animals with lower SNR, and reduces classification variability. The proposed method outperforms the state-of-the-art, demonstrating the value of optimizing both preprocessing and training for more reliable neural signal decoding.

References

[1] V. Sivaji, D. W. Grasse, S. A. Hays, J. E. Bucksot, R. Saini, M. P. Kilgard, and R. L. Rennaker II, "Restore: A wireless peripheral nerve stimulation system," *Journal of neuroscience methods*, vol. 320, pp. 26–36, 2019.

[2] T. Denison and M. J. Morrell, "Neuromodulation in 2035: the neurology future forecasting series," *Neurology*, vol. 98, no. 2, pp. 65–72, 2022.

[3] A. Coviello, F. Linsalata, U. Spagnolini, and M. Magarini, "Artificial neural networks-based real-time classification of eng signals for implanted nerve interfaces," *IEEE Journal on Selected Areas in Communications*, 2024.

[4] C. Silveira, R. Khushaba, E. Brunton, and K. Nazarpour, "Spatio-temporal feature extraction in sensory electroneurographic signals," *Philosophical Transactions of the Royal Society A*, vol. 380, no. 2228, p. 20210268, 2022.

[5] T. R. Farrell and R. F. Weir, "The optimal controller delay for myoelectric prostheses," *IEEE Transactions on neural systems and rehabilitation engineering*, vol. 15, no. 1, pp. 111–118, 2007.

[6] L. Citi, J. Carpaneto, K. Yoshida, K.-P. Hoffmann, K. P. Koch, P. Dario, and S. Micera, "On the use of wavelet denoising and spike sorting techniques to process electroneurographic signals recorded using intraneural electrodes," *Journal of neuroscience methods*, vol. 172, no. 2, pp. 294–302, 2008.

[7] A. Coviello, R. S. Augusto, D. Gottardello, M. Esposito, M. Manetti, U. Spagnolini, and M. Magarini, "Strategies for efficient eng signal classification using data augmentation and data balancing techniques," in *2024 IEEE International Conference on Metrology for eXtended Reality, Artificial Intelligence and Neural Engineering (MetroXRAINE)*. IEEE, 2024, pp. 576–581.

[8] D. Zhou, Q. Xu, J. Wang, H. Xu, L. Kettunen, Z. Chang, and F. Cong, "Alleviating class imbalance problem in automatic sleep stage classification," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–12, 2022.

[9] E. Brunton, C. W. Blau, and K. Nazarpour, "ENG recordings in response to mechanical stimulation," 2017. [Online]. Available: https://data.ncl.ac.uk/articles/dataset/ENG_recordings_in_response_to_mechanical_stimulation/10280846

[10] Y. He, F. Corradi, C. Shi, S. van der Ven, M. Timmermans, J. Stuijt, P. Detterer, P. Harpe, L. Lindeboom, E. Hermeling *et al.*, "An implantable neuromorphic sensing system featuring near-sensor computation and send-on-delta transmission for wireless neural sensing of peripheral nerves," *IEEE Journal of Solid-State Circuits*, vol. 57, no. 10, pp. 3058–3070, 2022.

[11] W. J. Freeman, M. D. Holmes, B. C. Burke, and S. Vanhatalo, "Spatial spectra of scalp eeg and emg from awake humans," *Clinical Neurophysiology*, vol. 114, no. 6, pp. 1053–1068, 2003.