SHAP-Guided LightGBM Classification of Neuropathic EMG Signals

Massimo Coppotelli[1], Aziz Gaaya[1], Patricia Conde-Cespedes[1], Behçet Uğur Töreyin[2], and Maria Trocan[1]

[1]Institut Supérieur d'Électronique de Paris (ISEP), France
[2]Istanbul Technical University, Turkey
massimo.coppotelli@isep.fr, aziz.gaaya@eleve.isep.fr, patricia.conde-cespedes@isep.fr, toreyin@itu.edu.tr, maria.trocan@isep.fr

Abstract. Accurate identification of neuropathies from electromyography (EMG) is crucial for automated diagnosis and future wearable screening systems. In this work, invasive EMG signals were processed to extract a 17 dimensional feature vector and classified as Healthy or Neuropathy using Light Gradient Boosting Machine (LightGBM). Model robustness was ensured through stratified 5 fold cross validation and repeated evaluation over 100 random dataset splits. SHapley Additive exPlanations (SHAP) were then applied to assess feature relevance and interpretability.

Based on the SHAP ranking, we introduced a SHAP-guided Iterative Feature Elimination (SHIFE) strategy, which removes features according to their estimated importance. This approach was compared with an unguided Iterative Feature Elimination (IFE) baseline that evaluates multiple feature combinations at each reduction step. Both methods improve performance with respect to the full feature set: IFE reduces the feature vector to 6 features and increases mean accuracy and AUC, while SHIFE reduces it to 7 features, preserving accuracy and improving AUC.

Keywords: EMG classification · Neuropathy diagnosis · Explainable AI · SHAP · LightGBM · Feature selection

1 Introduction

Neuromuscular disorders are a broad family of conditions that disrupt the normal communication between nerves and muscles [1], often leading to progressive weakness, impaired motor control and long term functional limitations [2]. Among them, neuropathies arise from abnormalities of peripheral nerves and alter the recruitment patterns and morphology of motor unit activity detectable in electromyographic signals [3]. Electromyography (EMG) remains a key diagnostic tool for identifying neuropathic alterations, but its interpretation requires expert knowledge and relies on the qualitative assessment of complex signal patterns [4]. Automated EMG analysis can support clinical decision making by providing objective and reproducible evaluations of neuromuscular function [5].

Machine learning has increasingly been applied to EMG based diagnosis, using either feature based classifiers or deep learning models that operate directly on the raw signal [6], [7]. Building on this line of work, our previous study [6] investigated the neuromuscular diseases classification task by implementing a complete EMG processing pipeline together with three different Machine Learning models, and conducting extensive tests to evaluate the impact of preprocessing parameters on diagnostic performance.

A more extensive analysis performed in the present work, using a strict cross validation protocol and repeated evaluations across multiple dataset splits, confirmed that the Healthy vs. Neuropathy task exhibits stable and reproducible behavior across conditions. This makes neuropathy detection a well suited target for investigating model explainability and feature relevance, which constitute the main focus of this study.

To perform the classification, we selected Light Gradient Boosting Machine (LightGBM), a model that demonstrated strong and stable performance under rigorous testing [8], [9]. This allows us to address a more ambitious goal: understanding how specific EMG features contribute to model predictions and how this information can be used to improve both interpretability and efficiency. Explainability is essential for clinical adoption because transparent models provide insights that clinicians can relate to physiological mechanisms [10]. At the same time, identifying the most informative features enables the construction of lighter models with reduced computational cost, a key requirement for future integration into wearable systems based on surface EMG acquisition.

This study therefore applies Shapley Additive Explanations (SHAP) to interpret the classifier and verify that it relies on physiologically meaningful properties of EMG signals when discriminating between healthy and neuropathic activity. By analyzing SHAP attribution patterns, we evaluate whether the model focuses on clinically relevant signal characteristics rather than spurious correlations. Building on this interpretability framework, we further exploit the global SHAP feature ranking to implement a SHAP guided Iterative Feature Elimination (SHIFE) procedure and compare it against a baseline Iterative Feature Elimination (IFE). This dual analysis enables both a physiological validation of the model and a systematic investigation of how feature reduction affects performance and stability.

The remainder of the paper is organized as follows: Section 2 describes the dataset and the preprocessing pipeline; Section 3 details the feature extraction process; Section 4 presents the LightGBM training strategy, including cross validation and stability analysis; Section 5 introduces the two feature selection approaches, IFE and SHIFE; Section 6 reports the experimental results; finally, Section 7 concludes the paper and outlines future research directions.

2 Dataset and Preprocessing

The experiments in this study were conducted using an open access dataset of invasive electromyography recordings [11]. The dataset contains needle EMG

signals acquired during isometric contractions of the biceps brachii and deltoid muscles. A total of 241 participants were recorded, distributed across three diagnostic groups: healthy controls, myopathy patients and neuropathy patients. In the present work, recordings from myopathic subjects were not used.

All signals were recorded using needle electrodes, with amplitudes measured in microvolts, a frequency range from 16 Hz to 5000 Hz and a sampling rate of 32768 Hz. The dataset includes recordings from both the biceps and deltoid muscles, enabling the analysis of diagnostic patterns across different anatomical sites.

The control group consists of 50 healthy subjects aged between 17 and 56 years. Each subject provided four recordings, corresponding to the left and right biceps brachii and left and right deltoid, yielding 200 healthy EMG samples in total. These recordings are balanced across the two muscles.

The neuropathy group includes 93 subjects diagnosed with conditions such as anterior horn cell disorders, multiple sclerosis or radicular damage. Most neuropathic subjects contributed a single EMG recording, resulting in a total of 100 neuropathic samples, also balanced between biceps and deltoid recordings.

For the classification task in this study, a subset of 100 healthy recordings was randomly selected so as to match the 100 neuropathic samples. During this selection, the natural balance between deltoid and biceps recordings was preserved. The resulting dataset therefore consists of 200 EMG signals: 100 healthy and 100 neuropathic, each class containing an equal number of deltoid and biceps brachii recordings.

Figure 1 provides four representative EMG recordings from the dataset, illustrating healthy and neuropathic activity for both the deltoid and the biceps brachii muscles.

To prepare the EMG signals for feature extraction and classification, we adopted a preprocessing pipeline based on standard practices commonly used for biosignals. This pipeline builds upon the methodology implemented in [6], but here it has been explored in greater depth through extended experimentation and parameter evaluation. The objective was to identify the most appropriate filtering configurations for invasive EMG while preserving diagnostically relevant information for neuropathy detection.
The processing steps are as follows:

- **Butterworth Band Pass Filtering**: A Butterworth band pass filter was applied to remove low and high frequency components that do not contribute to diagnostic information. Low frequency artifacts typically originate from motion, baseline drift or electrode instability, while high frequency components are often dominated by noise. Filtering below 20 Hz helps reduce motion artifacts without removing relevant physiological events [12]. Although many EMG studies restrict the upper cutoff to 450 to 500 Hz [13], experiments confirmed that preserving higher frequency components can occasionally improve classification performance in our case.
- **Notch Filtering**: A 50 Hz notch filter was applied to suppress power line interference. Power line noise in EMG recordings often exhibits amplitudes

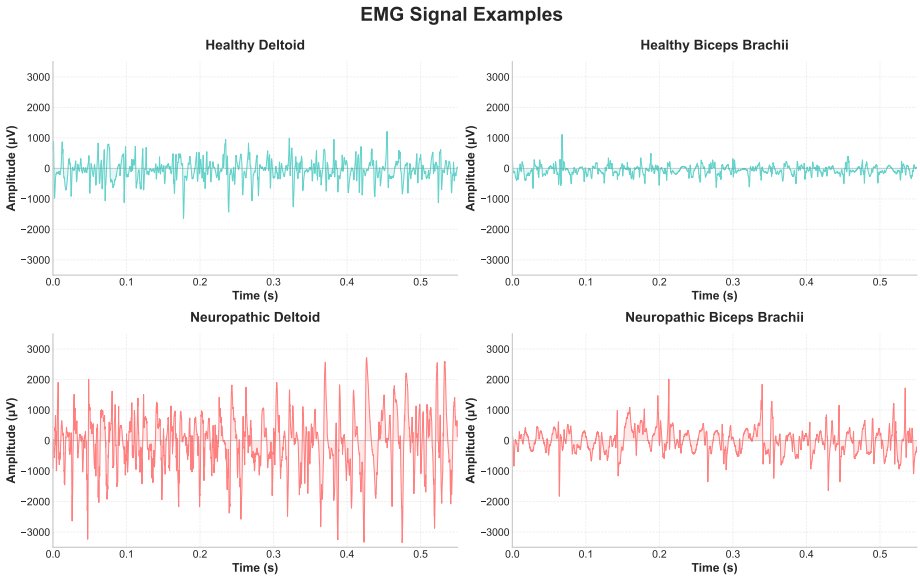

Fig. 1: Representative EMG recordings from the dataset, showing healthy and neuropathic signals for the deltoid and biceps brachii muscles.

that exceed the underlying muscle activity and can substantially distort both temporal and spectral features [14]. Removing this component ensures a cleaner representation of motor unit activity.

– **Resampling**: The original sampling rate of 32768 Hz provides a high resolution representation of EMG waveforms but introduces significant data redundancy for classification tasks. The signals were therefore resampled to lower rates. This step reduces computational cost and memory requirements without sacrificing essential information, an important consideration for future deployment on wearable systems.

After extensive cross validation and parameter exploration, the final signal processing configuration selected for this study consists of an eight fold downsampling (resulting in a 4096 Hz sampling rate), a third order Butterworth band pass filter between 20 and 2000 Hz and a 50 Hz notch filter.

3 Feature Extraction

Since LightGBM is a tree based model that operates effectively on structured tabular data, each EMG recording was transformed into a handcrafted feature vector. This approach follows the strategy adopted in our previous study and draws inspiration from the feature set introduced in [7], expanding it to capture a broader range of temporal, spectral, and morphological characteristics relevant to neuropathy detection.

The resulting feature vector is composed of 17 features. The first 16 describe properties of the EMG waveform, grouped into amplitude related, frequency based, variability related, distributional, and complexity based descriptors, and were already implemented in our previous study [6]. Here, an additional seventeenth binary feature, *Muscle*, was introduced to indicate whether the recording originates from the deltoid (0) or the biceps brachii (1). Exploratory analysis showed systematic distributional differences in several features between the two muscles; providing this information explicitly allows the classifier to account for anatomical variability without reducing the dataset size.

3.1 Feature Definitions

Let the EMG signal be represented as a sequence of N amplitude samples:

$$x = \{x_1, x_2, \ldots, x_N\}.$$

Its Power Spectral Density (PSD) is computed using Welch's method, which produces a set of discrete frequency bins $\{f_k\}_{k=1}^{K}$ and their corresponding spectral power values $\{P_k\}_{k=1}^{K}$. Here, k indexes the discrete frequency components of the PSD, and K denotes the total number of frequency bins returned by the Welch analysis.

Based on this representation, we constructed a feature vector composed of the following features.

1. Mean Absolute Value (MAV) : Represents the average absolute value of the signal. It is given by:

$$\mathrm{MAV} = \frac{1}{N}\sum_{i=1}^{N}|x_i|.$$

2. Root Mean Square (RMS) : Represents the square root of the mean of the squared signal values:

$$\mathrm{RMS} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}x_i^2}.$$

3. Zero Crossings (ZC) : The number of times the signal crosses zero amplitude

$$\mathrm{ZC} = \sum_{i=1}^{N-1}\mathbb{1}\left((x_i \cdot x_{i+1}) < 0\right),$$

where $\mathbb{1}$ is the indicator function, which is 1 if the condition is true and 0 otherwise.

4. Peak Frequency (f_{peak}) : The frequency corresponding to the highest power in the Power Spectral Density (PSD), calculated as:

$$f_{\mathrm{peak}} = f_{\arg\max_k(P_k)}.$$

5. Mean Frequency (f_{mean}) : The weighted average of all frequencies in the signal, based on their power:

$$f_{\text{mean}} = \frac{\sum_k f_k P_k}{\sum_k P_k}.$$

6. Median Frequency (f_{med}) : The frequency that divides the cumulative power spectrum into two equal halves. Find the smallest m such that:

$$\sum_{k=1}^{m} P_k \geq \frac{1}{2} \sum_{k=1}^{K} P_k, \quad \text{then} \quad f_{\text{med}} = f_m.$$

7. Total Power (TP) : The total energy contained in the frequency spectrum:

$$\text{TP} = \sum_{k=1}^{K} P_k.$$

8. Variance (σ^2) : The spread of the signal's amplitude values:

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^2, \qquad \mu = \frac{1}{N} \sum_{i=1}^{N} x_i,$$

where μ is the mean amplitude of the EMG recording.

9. Shannon Entropy (S) : A measure of randomness or complexity in the PSD. Let $p_k = P_k / \sum_j P_j$, then:

$$S = -\sum_{k=1}^{K} p_k \log_2(p_k).$$

10. Skewness (Sk) : Measures the asymmetry of the signal's amplitude distribution:

$$Sk = \frac{\frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^3}{\sigma^3},$$

where σ is the standard deviation of the signal's amplitude distribution.

11. Kurtosis (K) : Measures the "tailedness" of the signal's amplitude distribution:

$$K = \frac{\frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^4}{\sigma^4} - 3.$$

12. Petrosian Fractal Dimension (PFD) : Quantifies the fractal complexity of the signal [15]:

$$\text{PFD} = \frac{\log_{10}(N)}{\log_{10}(N) + \log_{10}\left(\frac{N}{N+0.4 \cdot ZC}\right)}.$$

13. Slope Sign Changes (SSC) : The number of changes in the slope's sign:

$$\mathrm{SSC} = \sum_{i=1}^{N-2} \mathbb{1}\left((x_{i+1} - x_i) \cdot (x_{i+2} - x_{i+1}) < 0\right).$$

14. Waveform Length (WL) : The cumulative length of the signal waveform:

$$\mathrm{WL} = \sum_{i=1}^{N-1} |x_{i+1} - x_i|.$$

15. Signal to Noise Ratio (SNR) : The ratio of signal power to noise power:

$$\mathrm{SNR} = 10 \log_{10}\left(\frac{\textit{signal power}}{\textit{noise power}}\right),$$

where *signal power* is the mean squared amplitude of the EMG signal, while *noise power* is the mean squared amplitude of the same signal after applying a stop band filter between 20 and 500 Hz, which suppresses the physiological EMG components and preserves primarily the low frequency and high frequency portions that are most likely to correspond to noise.

16. Recurrence (R) : The proportion of PSD values exceeding the mean plus one standard deviation:

$$R = \frac{1}{K} \sum_{k=1}^{K} \mathbb{1}\left(P_k > \mu_P + \sigma_P\right),$$

where μ_P, σ_P, and K are the mean, standard deviation, and length of the PSD.

17. Muscle (M) : Indicator for which muscle the EMG signal was acquired from:

$$\mathrm{M} = \begin{cases} 0, & \text{deltoid} \\ 1, & \text{biceps brachii} \end{cases}.$$

Together, these 17 features provide a structured numerical representation of each EMG recording, capturing its amplitude, frequency content, variability, morphological complexity, and muscle-specific context. This feature vector serves as the input to the LightGBM classifier described in the next section.

4 LightGBM Training Strategy

Light Gradient Boosting Machine (LightGBM) is a gradient boosting framework based on decision trees, optimized for efficiency and scalability [8]. It constructs an ensemble of weak learners, where each tree corrects the residual errors of the previous ones by minimizing a differentiable loss function.

To train and evaluate the model, the dataset was first divided into an 80% training plus validation set and a 20% test set, with the test set kept entirely unseen during hyperparameter tuning. The split was fully balanced across diagnostic class and muscle type, ensuring that healthy deltoid, healthy biceps, neuropathy deltoid, and neuropathy biceps each accounted for 25% of the samples in both partitions.

The 80% training subset was used for a stratified 5-fold cross validation, again preserving the four-way balance in every fold. For each candidate configuration in the hyperparameter grid, LightGBM was trained on four folds and validated on the remaining one. The resulting performance metrics were averaged across folds to identify the best set of hyperparameters. The hyperparameter grid explored variations in learning rate, tree complexity, sample and feature subsampling ratios, and bagging frequency, allowing the model to balance expressive power and generalization.

Given the relatively small size of the dataset, we observed that some hyperparameter configurations appearing optimal during cross validation could still exhibit instability when trained and tested on different dataset splits. For this reason, once the best hyperparameters were selected, they were fixed and the final training and testing phases were repeated 100 times using different random train–test partitions. This procedure provided mean performance metrics and their standard deviations, allowing us to quantitatively assess the stability and robustness of the resulting model.

Model evaluation relied on four metrics: accuracy, sensitivity, specificity, and AUC (Area Under the Curve). Accuracy measures overall correctness; sensitivity quantifies the ability to detect neuropathy; specificity measures the ability to identify healthy recordings; and AUC provides a threshold-independent measure of class separability.

5 Feature Selection

Feature selection helps reduce the dimensionality of the EMG feature space, improving efficiency and clarifying which signal characteristics are most relevant for neuropathy detection. Given the heterogeneous nature of the 17 dimensional feature vector, identifying the most informative subset is essential both for lightweight model design and for verifying that the classifier relies on physiologically meaningful features. To address this, we complemented the SHAP-based interpretability analysis with two feature selection strategies: a baseline Iterative Feature Elimination (IFE) procedure and a SHAP-Guided Iterative Feature Elimination (SHIFE) approach.

Importantly, in both procedures every occurrence of "train LightGBM" refers to the complete model selection pipeline described in Section 4, including 5-fold cross validation, hyperparameter grid search, and final evaluation repeated over 100 random train–test splits to obtain mean performance estimates and assess the model's stability.

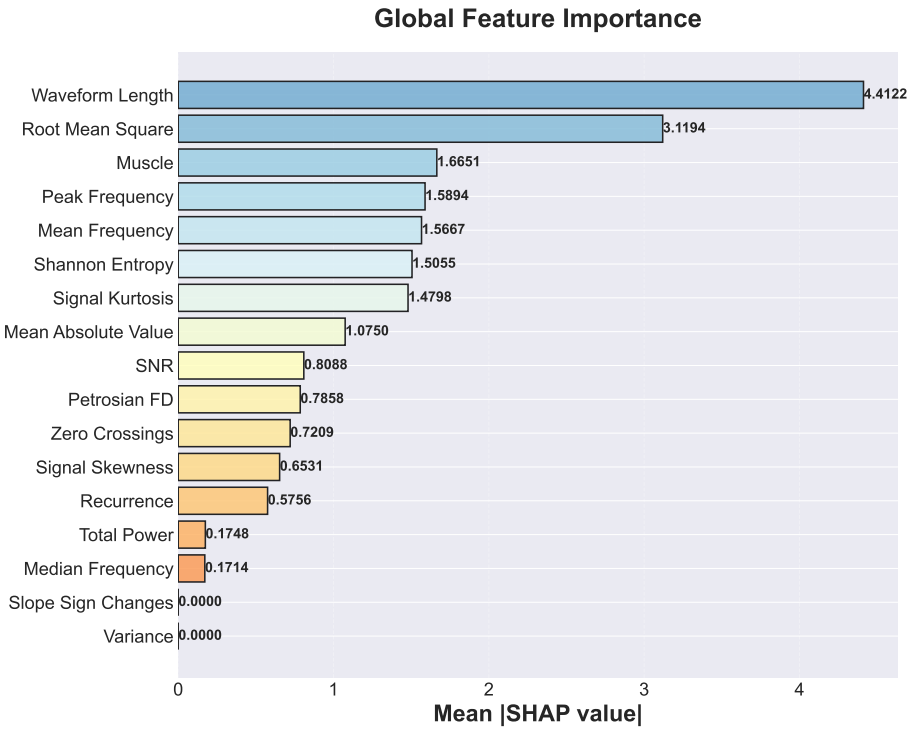

Fig. 2: SHAP summary plot highlighting feature importance.

5.1 Iterative Feature Elimination (IFE)

The IFE method evaluates feature subsets in a hierarchical way, starting from all 17 features and progressively reducing dimensionality by retaining, at each step, the subset that achieves the best predictive performance. Subsets of decreasing size are exhaustively evaluated until reaching a single-feature model. The full procedure is summarized in Figure 3 (left).

This method results in 153 total iterations, which corresponds to evaluating every step of the hierarchical reduction from 17 to 1 feature. Although computationally expensive, IFE provides a strong baseline for feature selection, enabling a direct comparison with the SHAP-guided approach described in the next two sections.

5.2 SHAP Overview

To investigate the contribution of each feature to the LightGBM decision process, we employed SHAP (SHapley Additive exPlanations), a unified framework for interpreting machine learning models based on cooperative game theory [16]. SHAP assigns an importance value to each feature by computing its contribution to the prediction relative to a reference baseline.

For a given sample x and model f, the SHAP value ϕ_i associated with feature i represents the marginal contribution of that feature averaged over all possible coalitions of features. Formally, the Shapley value for feature i is defined as:

$$\phi_i(f,x) = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!\,(|F|-|S|-1)!}{|F|!} \left[f(x_{S\cup\{i\}}) - f(x_S) \right],$$

where F is the full set of features, S is a subset of features not containing i, and x_S is the input restricted to the features in S.

In practice, SHAP values quantify how much each feature pushes the prediction toward the neuropathy class or toward the healthy class, and their absolute magnitude reflects feature importance. SHAP values can be positive or negative: positive values indicate that the feature increases the model's confidence toward the neuropathy class, while negative values push the prediction toward the healthy class. Values near zero indicate negligible influence on the specific prediction.

A global SHAP importance ranking (Fig. 2) was computed using the LightGBM model trained on the full set of 17 features. For each feature i, the global importance was quantified as the mean absolute SHAP value across all samples:

$$I_i = \frac{1}{N} \sum_{j=1}^{N} |\phi_i(x_j)|.$$

This ranking defines a consistent model-aware ordering of features from most to least influential and serves as the basis for the iterative reduction of the feature vector explained in the next section.

5.3 SHAP-Guided Iterative Feature Elimination (SHIFE)

SHIFE integrates model interpretability into the selection process. First, a LightGBM model is trained on all 17 features and global SHAP importances are computed. Features are then sorted by decreasing importance, and progressively reduced by removing the least influential ones. At each dimensionality level, the full LightGBM training pipeline is executed to assess performance and stability. The procedure is shown in Figure 3 (right).

A global SHAP feature importance plot is reported in Fig. 2, and the pseudocode for both IFE and SHIFE is provided in Fig. 3.

6 Experimental Framework

For Iterative Feature Elimination (IFE) baseline, features were removed in the following order: ZC, MAV, SNR, f_{peak}, σ^2, RMS, R, SSC, S, Sk, f_{mean}, f_{med}, WL, K, PFD, M, leaving Total Power (TP) as the final single feature. The best subset identified by IFE contains 6 features:

$$[f_{med}, \mathrm{WL}, K, \mathrm{PFD}, \mathrm{M}, \mathrm{TP}],$$

Algorithm 1: IFE

$\mathcal{F} \leftarrow$ full set of 17 features
$\mathcal{R} \leftarrow \emptyset$ ▷ Results storage
for $k = 17$ **downto** 1 **do**
  $\mathcal{F}_{\text{best}} \leftarrow$ None
  $m_{\text{best}} \leftarrow -\infty$
  for each subset $\mathcal{S}$ of size k from $\mathcal{F}$ **do**
    Train LightGBM on $\mathcal{S}$
    Compute metric m
    if $m > m_{\text{best}}$ **then**
      $m_{\text{best}} \leftarrow m$
      $\mathcal{F}_{\text{best}} \leftarrow \mathcal{S}$
    end if
  end for
  $\mathcal{F} \leftarrow \mathcal{F}_{\text{best}}$
  $\mathcal{R}[k] \leftarrow (\mathcal{F}_{\text{best}}, m_{\text{best}})$
end for
return best feature set from $\mathcal{R}$

Algorithm 2: SHIFE

$\mathcal{F} \leftarrow$ full set of 17 features
Train LightGBM on $\mathcal{F}$
Compute SHAP importance
Sort $\mathcal{F}$ by importance (descending)
$\mathcal{R} \leftarrow \emptyset$ ▷ Results storage
Store metric for full set in $\mathcal{R}[17]$
for $k = 16$ **downto** 1 **do**
  $\mathcal{F}_k \leftarrow$ first k features of $\mathcal{F}$
  Train LightGBM on $\mathcal{F}_k$
  Compute metric m
  $\mathcal{R}[k] \leftarrow (\mathcal{F}_k, m)$
end for
return best feature set from $\mathcal{R}$

Fig. 3: Pseudocode of the Iterative Feature Elimination (IFE, left) and SHAP-Guided Iterative Feature Elimination (SHIFE, right) procedures.

yielding 0.93 ± 0.03 Accuracy, 0.91 ± 0.06 Sensitivity, 0.94 ± 0.05 Specificity, and 0.97 ± 0.02 AUC.

For the SHAP-Guided Iterative Feature Elimination (SHIFE) method, features were removed in the sequence: σ^2, SSC, f_{med}, TP, R, Sk, ZC, PFD, SNR, MAV, K, S, f_{mean}, f_{peak}, M, RMS, leaving Waveform Length (WL) as the final feature. The best SHIFE subset includes 7 features:

$$[K, S, f_{mean}, f_{peak}, \text{M}, \text{RMS}, \text{WL}],$$

achieving 0.92 ± 0.04 Accuracy, 0.90 ± 0.06 Sensitivity, 0.94 ± 0.05 Specificity, and 0.97 ± 0.02 AUC.

The comparison among the full feature vector, the best IFE subset, and the best SHIFE subset is summarized in Table 1. Figure 4 illustrates how the four evaluation metrics evolve as a function of feature vector dimensionality, comparing IFE and SHIFE. The metric values are obtained over 100 train–test repetitions using different dataset splittings.

Table 1: Comparison between full and reduced feature sets.

Feature Set	Num Features	Accuracy	Sensitivity	Specificity	AUC
Full feature vector	17	0.92 ± 0.03	0.90 ± 0.06	0.94 ± 0.05	0.96 ± 0.03
Best IFE subset	6	0.93 ± 0.03	0.91 ± 0.06	0.94 ± 0.05	0.97 ± 0.02
Best SHIFE subset	7	0.92 ± 0.04	0.90 ± 0.06	0.94 ± 0.05	0.97 ± 0.02

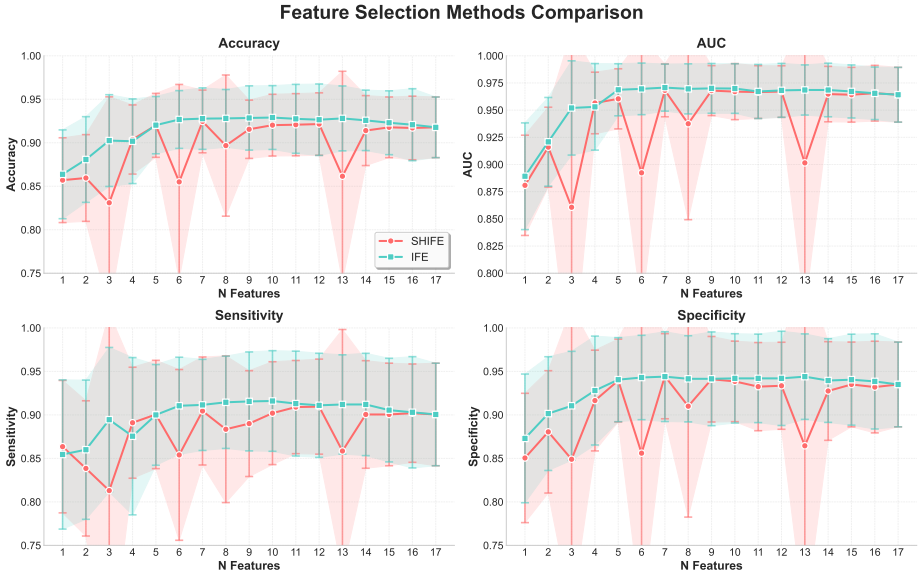

Fig. 4: Evolution of Accuracy, Sensitivity, Specificity, and AUC as a function of feature vector dimensionality for both IFE and SHIFE.

The SHAP analysis provides valuable insight into how the LightGBM classifier uses the EMG feature space to discriminate between healthy and neuropathic signals. The highest ranked features, Waveform Length (WL) and Root Mean Square (RMS), are both strongly related to the overall amplitude of the EMG signal. This observation is physiologically meaningful, since neuropathies are known to produce motor unit action potentials with increased amplitude [17]. The dominance of amplitude related features in the SHAP ranking therefore indicates that the model relies on clinically relevant characteristics rather than on spurious correlations.

Another interesting result is the high SHAP importance of the Muscle (M) feature, a binary indicator distinguishing recordings from the deltoid and biceps brachii. Although this feature carries no diagnostic information on its own, its interaction with the remaining feature set appears highly informative. Several EMG descriptors differ systematically between the two muscles even for signals originating from the same individual and having the same health status. Including this indicator allows the classifier to contextualize the other features and prevents muscle dependent variability from being misinterpreted as diagnostic variation.

Regarding feature selection, both methods demonstrate that the full 17 dimensional feature vector is not necessary to achieve high performance. The IFE method reduces the vector to 6 features while improving average Accuracy, AUC, and Sensitivity by 1%. The SHIFE method yields a 7 feature subset that increases the AUC by 1%. Although the two subsets overlap on only three features (WL,

M and K), this convergence indicates that amplitude driven descriptors and muscle context are systematically selected as core predictors by both methods. These reductions decrease the computational cost of feature extraction, model training, and inference, which is particularly relevant for future deployment in wearable EMG based systems.

Comparing the two selection strategies, Figure 4 shows that IFE provides a more stable and consistently superior performance across all feature vector sizes, while SHIFE exhibits more irregular trends and larger standard deviations across the random splits, indicating greater sensitivity to dataset variability.

$\sum_{f=2}^{F} \binom{f}{f-1} = O(F^2)$, so IFE has quadratic growth in F, with each evaluated subset incurring the cost of cross validation and repeated training. SHIFE evaluates exactly one subset per size (after a single SHAP pass) and therefore grows only linearly, $O(F)$, plus the one-time SHAP computation. Both procedures are far more efficient than an exhaustive search over all nonempty subsets, whose cost is $\sum_{f=1}^{F} \binom{F}{f} = O(2^F)$.

Overall, these findings highlight that a compact and physiologically meaningful feature set is sufficient for accurate neuropathy classification and that SHAP-based explainability can guide the design of lightweight models suitable for wearable EMG based diagnostic tools.

7 Conclusion

This study investigated the use of LightGBM for the classification of healthy vs. neuropathic EMG signals, with a particular focus on explainability and feature selection. After establishing a rigorous and balanced cross validation framework, the model demonstrated high accuracy, sensitivity, specificity, and AUC, confirming that neuropathy detection is a stable and reproducible task under the proposed pipeline. SHAP analysis revealed that the classifier relies primarily on physiologically meaningful features, particularly those related to EMG amplitude, and highlighted the importance of contextual information such as the muscle of origin.

Two feature selection strategies were evaluated: an Iterative Feature Elimination approach (IFE) based on systematic subset evaluation and a SHAP guided Iterative Feature Elimination (SHIFE). Both methods produced compact feature vectors while maintaining or improving performance relative to the full 17 feature set. These findings show that accurate and efficient neuropathy classification can be achieved using a small and interpretable subset of EMG features, improving both the feasibility of deployment in wearable systems and the transparency of the decision making process.

Future work may extend this framework in several directions. An important step is the validation of the proposed pipeline on non invasive surface EMG recordings, moving toward the development of wearable and fully automated neuromuscular screening tools. Additionally, future studies could investigate hybrid models that combine interpretable feature based classifiers with deep neural networks. While the former provide transparent decision making, deep models

can extract subtle temporal and spectral patterns directly from raw EMG signals that may not be fully captured by handcrafted descriptors. Integrating these two approaches may further enhance diagnostic performance while maintaining clinical interpretability.

Acknowledgments. The authors would like to acknowledge the creators of the publicly available electromyographic dataset used in this study [11].

References

1. B. M. Morrison, “Neuromuscular diseases,” Seminars in Neurology, vol. 36, no. 5, pp. 409–418, Oct. 2016.DOI: 10.1055/s-0036-1586263.
2. J. Deenen, C. Horlings, J. Verschuuren, A. Verbeek, and B. van Engelen, “The epidemiology of neuromuscular disorders: A comprehensive overview of the literature,” Journal of Neuromuscular Diseases, vol. 2, no. 1, pp. 73–85, 2015. DOI: 10.3233/JND-140045.
3. Devon I. Rubin, Christopher J. Lamb, Chapter 3 - The role of electrodiagnosis in focal neuropathies, Editor(s): Colin Chalk, Handbook of Clinical Neurology, Elsevier, Volume 201, 2024, Pages 43-59, ISSN 0072-9752, ISBN 9780323901086, https://doi.org/10.1016/B978-0-323-90108-6.00010-7.
4. C. Denby and T. Stone, “Chapter 18- Neurological measurement,” in Clinical Engineering, 2nd ed., A. Taktak, P. S. Ganney, D. Long, and R. G. Axell, Eds., Academic Press, 2020, pp. 309–319. DOI: 10.1016/B978-0-08-102694-6.00018-8.
5. M. A. Taha and J. A. Morren, “The role of artificial intelligence in electrodiagnostic and neuromuscular medicine: Current state and future directions,” Muscle&Nerve, vol. 69, no. 3, pp. 260-272, Dec. 2023. DOI: 10.1002/mus.28023.
6. M. Coppotelli, D. Albert, J. G. Barrena, R. E. Khoury, P. Conde-Cespedes and M. Trocan, "Machine Learning Models for EMG-Based Diagnosis of Neuromuscular Disorders," 2025 20th International Conference on PhD Research in Microelectronics and Electronics (PRIME), Taormina, Italy, 2025, pp. 1-4, doi: 10.1109/PRIME66228.2025.11203505.
7. N. Sadi-Ahmed, A. Messekher, S. Namane, M. Kedir-Talha, and S. Mekaoui, “Automatic diagnosis of neuromuscular diseases from electromyographic (EMG) records,” in Proc. 2017 5th Int. Conf. Electrical Engineering- Boumerdes (ICEE-B), Boumerdes, Algeria, 2017, pp. 1–6. DOI: 10.1109/ICEE-B.2017.8192195.
8. G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, “LightGBM: A highly efficient gradient boosting decision tree,” in Advances in Neural Information Processing Systems (NeurIPS), vol. 30, 2017.
9. P. Florek and A. Zagdański, "Benchmarking state-of-the-art gradient boosting algorithms for classification," DOI: 10.48550/arXiv.2305.17094
10. Finzel B. Current methods in explainable artificial intelligence and future prospects for integrative physiology. Pflugers Arch. 2025 Apr;477(4):513-529. doi: 10.1007/s00424-025-03067-7. Epub 2025 Feb 25. PMID: 39994035; PMCID: PMC11958383.
11. A. Messekher, S. Mekaooui, M. Kedir-Talha, M. I. Kediha and F. Mostefaoui, Electromyographic data recordings during isometric contractions of the biceps brachii and deltoid collected from patients and healthy subjects, version V1, Mendeley Data, 2023. DOI: 10.17632/543xpjycj9.1.

12. M. Boyer, L. Bouyer, J.-S. Roy, and A. Campeau-Lecours, “Reducing noise, artifacts and interference in single-channel EMG signals: A review,” Sensors (Basel, Switzerland), vol. 23, no. 6, p. 2927, 2023. DOI: 10.3390/s23062927.
13. R. G. T. Mello, L. F. Oliveira, and J. Nadal, “Digital Butterworth filter for subtracting noise from low magnitude surface electromyogram,”Computer Methods and Programs in Biomedicine, vol. 87, no. 1, pp. 28–35, 2007. DOI: 10.1016/j.cmpb.2007.04.004.
14. J. Piskorowski, “Powerline interference rejection from sEMG signal using notch filter with transient suppression,” in Proc. 2012 IEEE Int. Instrumentation and Measurement Technology Conf. (I2MTC), 2012, pp. 1447–1451. DOI: 10.1109/I2MTC.2012.6229332.
15. A. Petrosian, "Kolmogorov complexity of finite sequences and recognition of different preictal EEG patterns," Proceedings Eighth IEEE Symposium on Computer-Based Medical Systems, Lubbock, TX, USA, 1995, pp. 212-217, doi: 10.1109/CBMS.1995.465426.
16. S. Lundberg, S. Lee, "A Unified Approach to Interpreting Model Predictions," DOI: 10.48550/arXiv.1705.07874.
17. H. S. Milner-Brown, R. B. Stein, R. G. Lee, "Contractile and electrical properties of human motor units in neuropathies and motor neurone disease," DOI: doi.org/10.1136/jnnp.37.6.670.