Machine Learning Models for EMG-Based Diagnosis of Neuromuscular Disorders

Massimo Coppotelli
Institut Superieur d'Electronique de Paris (ISEP)
Paris, France
coppotelli.1705325@studenti.uniroma1.it

David Albert
Institut Superieur d'Electronique de Paris (ISEP)
Paris, France
david.albert@eleve.isep.fr

Julio García Barrena
Institut Superieur d'Electronique de Paris (ISEP)
Paris, France
julio.garcia-barrena@eleve.isep.fr

Romy El Khoury
Institut Superieur d'Electronique de Paris (ISEP)
Paris, France
romy.el-khoury@eleve.isep.fr

Patricia Conde-Cespedes
Institut Superieur d'Electronique de Paris (ISEP)
Paris, France
patricia.conde-cespedes@isep.fr

Maria Trocan
Institut Superieur d'Electronique de Paris (ISEP)
Paris, France
maria.trocan@isep.fr

***Abstract*—Neuromuscular disorders (NMDs), such as myopathies and neuropathies, affect the communication between nerves and muscles and often lead to serious functional impairments. This study investigates the impact of preprocessing methods and machine learning models on the automated classification of NMDs using electromyography (EMG) data. We compare Convolutional Neural Networks, Extreme Gradient Boosting, and Light Gradient Boosting Machine on two binary classification tasks: Healthy vs. Myopathy and Healthy vs. Neuropathy. The analysis is conducted on an open dataset of invasive needle EMG recordings, serving as a preliminary step toward the future use of non-invasive surface EMG in wearable diagnostic tools. The results highlight the importance of preprocessing and show promising performance across all models.**

***Index Terms*—EMG, Myopathy, Neuropathy, CNN, XGBoost, LightGBM, Binary Classification, Signal Processing**

I. Introduction

Neuromuscular disorders (NMDs) impair the communication between nerves and muscles, often leading to motor dysfunction and long-term complications [1]. Early and accurate diagnosis is critical to improve patient outcomes, but it requires specialized interpretation of diagnostic tools like electromyography (EMG) [2].

This study explores how different preprocessing strategies and machine learning models affect classification accuracy on EMG data for NMD diagnosis. Specifically, we compare Convolutional Neural Networks (CNNs), Extreme Gradient Boosting (XGBoost), and Light Gradient Boosting Machine (LightGBM) on two binary classification tasks: Healthy vs. Myopathy and Healthy vs. Neuropathy.

Although the dataset used in this study is based on invasive needle EMG recordings, the work serves as a preliminary step toward developing diagnostic tools based on non-invasive surface EMG. Due to the scarcity of publicly available non-invasive datasets for NMDs, understanding model behavior on high-quality invasive signals is essential before transitioning to real-world, accessible solutions.

II. Dataset Description

We use an open-access dataset [3] containing 5-second invasive EMG recordings collected via needle electrodes from 240 participants: 50 healthy controls, 97 myopathy patients, and 93 neuropathy patients. Recordings were taken from the biceps brachii and deltoid muscles, sampled at 32,768 Hz with a frequency range of 16–5000 Hz.

Each healthy subject provided four recordings, while most patients contributed one. To reduce class imbalance, we retained one random recording per muscle per healthy subject. The final dataset contains 100 Healthy, 97 Myopathy, and 100 Neuropathy recordings.

To ensure independence and minimize information leakage, all recordings from a given individual were kept within a single dataset partition, using the unique subject IDs embedded in file names. The dataset was split into 80% training, 10% validation, and 10% test sets. This process was repeated with different random seeds, and results were consistent across runs.

III. Data Preprocessing

A. Filtering, Resampling, and Subsampling

We applied a stardardized preprocessing pipeline designed to reduce noise and preserve diagnostically relevant features of the EMG signal [4]. A Butterworth band-pass filter was tested with low-cut frequencies of 20 Hz or none, and high-cut frequencies ranging from 500 to 4000 Hz or none, depending on the configuration [5]. A 50 Hz notch filter was consistently applied to suppress power line interference [6].

Frequency resampling was performed (with resampling factors ranging from 1 to 16) to reduce redundancy while retaining essential frequency information. EMG signals were then segmented into overlapping subsamples of various durations (from 0.005 to 5 seconds). The subsample overlap was fixed

979-8-3315-0390-1/25/$31.00 © 2025 IEEE

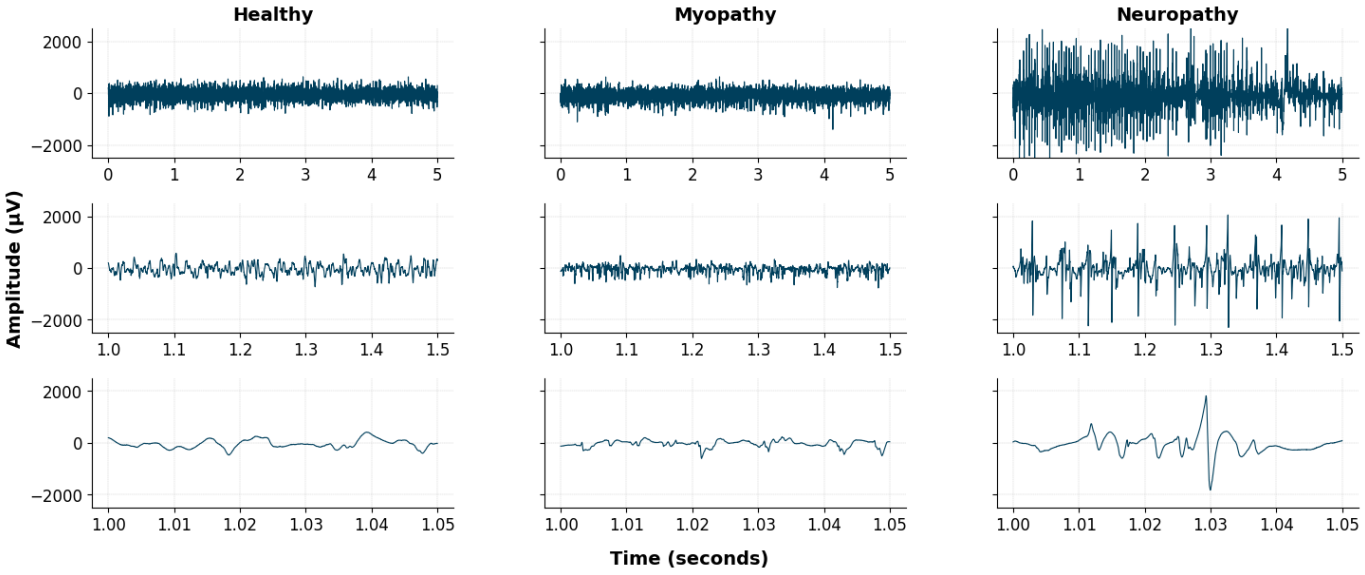

Fig. 1. Sample plots of EMG recordings from the dataset. The first row displays full 5-second EMG recordings for the Healthy (first column), Myopathy (second column), and Neuropathy (third column) classes, with time on the x-axis (in seconds) and amplitude on the y-axis (in microvolts). The second and third rows show subsamples of these recordings: the second row contains 0.5-second subsamples, and the third row contains 0.05-second subsamples.

at 30%, as preliminary tests showed minimal performance variation with different overlap values.

We evaluated more than 150 preprocessing configurations, combining filtering, resampling, and subsampling parameters. Each was tested across all models to determine optimal setups for both classification tasks.

B. Input Formats

Since CNNs are particularly effective at learning spatial hierarchies and patterns directly from raw data [7], the EMG subsamples were fed directly into the network as input. On the other hand, XGBoost and LightGBM are gradient boosting models that excel when working with structured data [8]. Therefore, for these models, a set of relevant features was extracted from each subsample, creating feature vectors that were used as input for classification.

C. Feature Vector Construction for XGBoost and LightGBM

Taking inspiration from [9], for XGBoost and LightGBM, each EMG subsample was converted into a 16-dimensional feature vector designed to capture a wide range of signal characteristics. These features were selected to reflect clinical and physiological properties typically altered in NMDs, and were grouped as follows:

- **Amplitude-related features**: Mean Absolute Value, Root Mean Square, and Total Power quantify signal strength and energy.
- **Frequency-based features**: Peak Frequency, Mean Frequency, and Median Frequency provide insight into spectral content, often altered in pathological EMG.
- **Variability and randomness**: Variance and Shannon Entropy assess the dispersion and complexity of the signal.
- **Morphological features**: Zero Crossings, Slope Sign Changes, and Waveform Length measure signal irregularity and oscillatory behavior.
- **Distribution features**: Skewness and Kurtosis evaluate asymmetry and sharpness of the amplitude distribution.
- **Complexity and clarity**: Petrosian Fractal Dimension reflects signal complexity, while Signal-to-Noise Ratio and Recurrence assess signal clarity and periodicity.

This comprehensive feature set was applied uniformly to both classification tasks. Future studies could explore task-specific feature selection to optimize performance and reduce computational cost.

IV. Classification Models

We evaluated three machine learning models: CNN [10], XGBoost [11], and LightGBM [12]. CNN were trained directly on raw EMG subsamples, while XGBoost and LightGBM used the 16-dimensional feature vectors described in Section III-C.

The CNN consisted of three 1D convolutional blocks followed by two fully connected layers and dropout for regularization. All models were trained using the same dataset partitions and preprocessing configurations described earlier. Model hyperparameters were selected based on preliminary experiments and kept fixed across all preprocessing configurations; full details are available upon request.

V. Results

Model performance was evaluated on both individual EMG subsamples and full recordings. While models were trained on subsamples, our goal is to assess a patient's condition based on an entire EMG recording. Therefore, we applied a majority voting strategy during testing: the predicted label for each

TABLE I
MODELS AND PREPROCESSING PERFORMANCE COMPARISON FOR 'HEALTHY VS. MYOPATHY'

Model	Majority Voting			Single Subsamples			Time (s)	Dataset Configuration			
	F1	Sens.	Spec.	F1	Sens.	Spec.		Freq (Hz)	Sub. Len. (s)	Low Cut (Hz)	High Cut (Hz)
CNN	0.96	0.92	1.00	0.75	0.75	0.74	0.9	32768	0.005	20	2000
CNN	0.88	0.85	0.92	0.77	0.79	0.70	0.7	32768	0.005	-	-
CNN	0.77	0.92	0.50	0.77	0.92	0.50	0.2	16384	5	20	-
CNN	0.59	0.62	0.50	0.59	0.62	0.50	0.1	32768	5	-	-
XGBoost	0.89	0.92	0.83	0.78	0.78	0.76	0.3	4096	0.1	20	2000
XGBoost	0.80	0.77	0.83	0.77	0.75	0.78	0.4	32768	0.05	-	-
XGBoost	0.88	0.85	0.92	0.88	0.85	0.92	0.1	8192	5	20	4000
XGBoost	0.70	0.62	0.83	0.70	0.62	0.83	0.1	32768	5	-	-
LightGBM	0.89	0.92	0.83	0.76	0.78	0.72	0.4	4096	0.1	20	2000
LightGBM	0.85	0.85	0.83	0.78	0.76	0.80	0.7	32768	0.05	-	-
LightGBM	0.88	0.85	0.92	0.88	0.85	0.92	0.1	16384	5	20	2000
LightGBM	0.67	0.62	0.75	0.67	0.62	0.75	0.1	32768	5	-	-

TABLE II
MODELS AND PREPROCESSING PERFORMANCE COMPARISON FOR 'HEALTHY VS. NEUROPATHY'

Model	Majority Voting			Single Subsamples			Time (s)	Dataset Configuration			
	F1	Sens.	Spec.	F1	Sens.	Spec.		Freq (Hz)	Sub. Len. (s)	Low Cut (Hz)	High Cut (Hz)
CNN	0.96	0.92	1.00	0.91	0.90	0.91	0.1	4096	0.2	20	2000
CNN	1.00	1.00	1.00	0.88	0.89	0.86	0.3	32768	0.01	-	-
CNN	0.96	0.92	1.00	0.96	0.92	1.00	0.1	4096	5	20	2000
CNN	0.83	0.83	0.83	0.83	0.83	0.83	0.1	32768	5	-	-
XGBoost	0.96	0.92	1.00	0.94	0.92	0.96	0.1	2048	1	20	500
XGBoost	1.00	1.00	1.00	0.96	0.96	0.96	0.1	32768	0.2	-	-
XGBoost	0.96	0.92	1.00	0.96	0.92	1.00	0.1	2048	5	20	500
XGBoost	0.96	0.92	1.00	0.96	0.92	1.00	0.1	32768	5	-	-
LightGBM	1.00	1.00	1.00	0.96	0.95	0.97	0.3	8192	0.2	-	4000
LightGBM	1.00	1.00	1.00	0.95	0.97	0.92	0.2	32768	0.2	-	-
LightGBM	1.00	1.00	1.00	1.00	1.00	1.00	0.1	2048	5	20	1000
LightGBM	1.00	1.00	1.00	1.00	1.00	1.00	0.1	32768	5	-	-

recording corresponds to the most frequent prediction across its subsamples. This approach better reflects real-world usage, where clinicians assess full EMG segments rather than isolated fragments.

Tables I and II summarize results for the Healthy vs. Myopathy and Healthy vs. Neuropathy classification tasks, respectively. For each model, we report F1 score, sensitivity, and specificity at both the subsample and majority-vote level, along with execution time and preprocessing parameters (sampling frequency, subsample duration, filtering choices). "Time (s)" refers to the average total time required to classify a single EMG recording, including: subsampling time (if applied), filtering time for each subsample (if applied), feature extraction time for each subsample (only for gradient boosting models), classification time for each subsample, and final majority voting.

Each table is organized into three blocks, one for each model (CNN, XGBoost, LightGBM), and four rows per block. These rows correspond to different dataset configurations: (1) fully preprocessed (filtered, resampled, and subsampled), (2) subsampled only, (3) full-length recordings (filtered/resampled), and (4) raw, unprocessed data. This structure enables direct comparison of models across varying preprocessing pipelines.

A. Healthy vs. Myopathy

For the Healthy vs. Myopathy task, the CNN achieved the highest F1 score (0.96) on the fully preprocessed dataset, using short (0.005 s) subsamples. In this configuration, it clearly outperformed XGBoost and LightGBM, which also reached strong F1 scores (0.89) with lower computational cost.

However, this advantage diminished outside the fully preprocessed setting. While the CNN remained the best model with subsampled-only data, it underperformed on unsubsampled signals, achieving the lowest scores among the three. This may be due to the model's shallow architecture, which struggles to process longer time windows, or because the key diagnostic features in EMG (e.g., individual Motor Unit Action Potentials [13]) are better captured in short, localized segments.

XGBoost and LightGBM also achieved their best results on fully preprocessed data and showed marked drops in performance on raw signals, confirming the necessity of filtering and temporal segmentation for effective classification.

Overall, CNNs excel when preprocessing and segmentation isolate relevant signal events, while gradient boosting models offer a solid balance between accuracy and efficiency.

B. Healthy vs. Neuropathy

For the Healthy vs. Neuropathy task, all models achieved excellent results, with several configurations reaching an F1 score of 1.00 under majority voting. LightGBM consistently delivered perfect accuracy across all preprocessing setups, including raw and unfiltered data, highlighting its robustness.

CNN and XGBoost also reached perfect accuracy, but only in subsampled configurations without filtering. Notably, unlike in the Myopathy task, the CNN performed well even with longer subsamples, resulting in shorter inference times and making it a competitive option in terms of both accuracy and efficiency.

Overall, this task appears to be less challenging than Myopathy detection, with high separability between healthy and neuropathic EMG patterns. All three models performed reliably, though LightGBM stood out for its stability across conditions. However, the clear class separation observed here does not preclude the use of machine learning, which offers a flexible framework suited for more complex tasks and supports the future transition to surface EMG data.

C. Summary of Results

All three models achieved high performance across both classification tasks, with differences depending on preprocessing. For Healthy vs. Myopathy, CNN yielded the highest F1 score on fully preprocessed data, while XGBoost and LightGBM offered strong performance with lower computational cost.

The Healthy vs. Neuropathy task proved easier: multiple models reached perfect accuracy, especially in subsampled settings. LightGBM was the most robust, performing consistently well across all preprocessing configurations, including raw data.

Subsampling consistently improved performance, whereas filtering was essential for Myopathy detection but sometimes detrimental for Neuropathy, possibly due to loss of informative high-frequency content.

A prior study on the same dataset [9] reported 91% and 97% accuracy for the two tasks using SVMs with Huang-Hilbert-derived features. While that work focused on feature selection and accuracy, our study broadens the analysis by comparing multiple model types under diverse preprocessing conditions, reporting both performance and computational efficiency.

VI. Conclusions and Future Works

This study compared machine learning models for classifying neuromuscular disorders from EMG signals, analyzing the effect of preprocessing strategies. For the Healthy vs. Myopathy task, CNN achieved the best performance (F1 = 0.96) on fully preprocessed data, while XGBoost and LightGBM performed well with shorter inference time. In the Healthy vs. Neuropathy task, all models performed excellently, with LightGBM achieving perfect accuracy across all configurations, including raw data.

Subsampling significantly improved classification in both tasks. Filtering was crucial for Myopathy detection but could reduce performance for Neuropathy, highlighting the need to adapt preprocessing to task-specific signal characteristics.

These findings confirm the feasibility of fast, accurate classification using needle EMG and support the transition toward non-invasive solutions. Future work will focus on acquiring high-quality surface EMG data, adapting models to its properties, enabling low-power deployment on wearable devices. In parallel, improving model explainability will support clinical trust and integration.

Acknowledgment

The authors would like to acknowledge the creators of the publicly available electromyographic dataset used in this study [3].

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] 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.

[3] 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.

[4] 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. [Online]. Available: https://doi.org/10.3390/s23062927.

[5] 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. [Online]. Available: https://doi.org/10.1016/j.cmpb.2007.04.004.

[6] 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. [Online]. Available: https://doi.org/10.1109/I2MTC.2012.6229332.

[7] F. Shaheen, B. Verma, and M. Asafuddoula, "Impact of automatic feature extraction in deep learning architecture," in *Proc. 2016 Int. Conf. Digital Image Computing: Techniques and Applications (DICTA)*, Gold Coast, QLD, Australia, 2016, pp. 1–8. DOI: 10.1109/DICTA.2016.7797053.

[8] A. T. Gavito, D. Klabjan, and J. Utke, "Gradient-boosted based structured and unstructured learning," 2023. arXiv:2302.14299 [cs.LG]. [Online]. Available: https://arxiv.org/abs/2302.14299.

[9] 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.

[10] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998. DOI: 10.1109/5.726791.

[11] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD)*, San Francisco, CA, USA, 2016, pp. 785–794. DOI: 10.1145/2939672.2939785.

[12] 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.

[13] U. K. Dhand, "Motor unit potential," in *Encyclopedia of the Neurological Sciences*, 2nd ed., M. J. Aminoff and R. B. Daroff, Eds. Oxford, UK: Academic Press, 2014, pp. 117–119. DOI: 10.1016/B978-0-12-385157-4.00534-0. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780123851574005340.