HD-MUNet: integrating artificial intelligence and high-density electromyography for motor unit number estimation

HD-MUNet: integrating artificial intelligence and high-density electromyography for motor unit number estimation

M. Gagliardi, A. Ortiz, A. Conte, D. Belvisi, G. Leodori, G. Fabbrini, G. Ferrazzano and T.M. Vieira

***Abstract*— *Objective:* In this study we developed and evaluated the performance of HD-MUNet: a novel method for motor unit number estimation (MUNE), integrating high-density surface electromyograms (HD-sEMG) with artificial intelligence (AI). *Methods:* We designed a dual-branch neural network to process raw M-waves elicited by electrical stimulation and their incremental changes through convolutional and recurrent layers, producing stepwise estimates of recruited motor units (MUs). Because no experimental gold standard exists for counting MUs, a physiologically realistic model of the medial gastrocnemius was used to simulate HD-sEMG. Simulations included variations in MUs population (20-50-100-150-200-250-300), subcutaneous tissue thickness (4-8 mm), signal-to-noise ratio (0-8-16 dB), and stimulation protocol (50-100-200 steps), yielding 5040 subjects. HD-MUNet performance was benchmarked against two state-of-the-art methods: Incremental MUNE and StairFit. *Results:* HD-MUNet achieved consistently low median relative errors (≤ 7%) across conditions with high test-retest reliability (Pearson r > 0.995), whereas StairFit and Incremental MUNE showed strong sensitivity to noise and stimulation scenarios, with median errors ranging from −2% to −195% and −20% to −66%, respectively. Of practical relevance, differently from the other two methods, HD-MUNet performed equally well for the three stimulation steps considered. We confirmed this result on a set of experimental data. *Conclusion:* By combining HD-sEMG with AI, HD-MUNet achieved robust and reliable MUNE across a broad range of simulated conditions, independently of stimulation protocol. *Significance:* This method has immediate potential to simplify MUNE protocols in both healthy and pathological conditions, with direct applicability to clinical neuromuscular disease diagnosis and monitoring.**

***Index Terms*—Artificial Intelligence, electromyography, HD-sEMG, MUNE.**

I. Introduction

MOTOR Unit Number Estimation (MUNE) refers to a set of electrophysiological techniques used to estimate the number of motor units (MUs) within a specific muscle [1], [2]. MUNE has received increasing attention in recent years due to its reliability in diagnosing neuromuscular diseases, monitoring motor unit loss, and evaluating treatment effects [2]. Assessing the size and number of MUs within a muscle is fundamental for understanding both its functional capacity and structural organization. Recently, several MUNE-based techniques have proven valuable as diagnostic and research tools. For example, tracking motor unit loss in subjects suffering from Amyotrophic Lateral Sclerosis (ALS) [3], Spinal Muscular Atrophy (SMA) [4] and Spinal Cord Injury [5]. However, existing MUNE methods only provide an estimation of the total number of MUs, which can be imprecise. This is particularly critical in contexts requiring high sensitivity in MU quantification, such as early disease diagnosis.

The fundamental principle of MUNE relies on the relationship between the number of excited MUs and EMG amplitude. It is based on the assumption that a series of progressively stronger stimuli delivered to the nerve produce a stepwise increase in the compound muscle action potential (CMAP) amplitude, due to the progressive excitation of additional MUs [6]. With traditional approaches, MUNE is estimated as the ratio between the maximal CMAP amplitude and the average amplitude of a single MU action potential (SMUP) [1], [2]. These methods assume that each recruited MU contributes roughly equally to the total CMAP. More recently, CMAP scan methods have been proposed as a significant advancement over traditional MUNE techniques. A CMAP scan involves recording a high resolution stimulus-response curve by delivering hundreds of stimuli with gradually increasing intensity, from subthreshold to supramaximal levels [7]. Unlike traditional methods that sample a small subset of MUs [6], [8], [9], CMAP scan methods, such as StairFit [5], aim to capture responses from the entire motor unit pool, potentially reducing sampling bias. MUNE is then obtained using advanced mathematical and statistical modeling, such as Bayesian inference [10], model-fitting approaches [11], or staircase function optimization [5]. However, like traditional methods, CMAP scans rely on single-channel recordings, being vulnerable to sampling errors in heterogeneous muscles and

Manuscript submitted for review on October 2, 2025. This study was carried out within the «BREAKing BONDS» project – funded by the Ministero dell'Università e della Ricerca – within the PRIN 2022 program (D.D. 104 - 02/02/2022). This manuscript reflects only the authors' views and opinions and the Ministry cannot be considered responsible for them.

M. Gagliardi, T.M. Vieira and A. Ortiz are with the Laboratory for Engineering of the Neuromuscular System (LISiN), Dipartimento di Elettronica e Telecomunicazioni, Politecnico di Torino and with the Polito^{BIO}Med Lab, Politecnico di Torino, Torino, Corso Duca degli Abruzzi 29, 10129 Italy (correspondence e-mail: taian.martins@polito.it).

A. Conte, D. Belvisi, G. Leodori, and G. Fabbrini are with Department of Human Neurosciences, Sapienza University, Roma, Italy and with IRCCS Neuromed, Italy.

G. Ferrazzano is with Department of Human Neurosciences, Sapienza University, Roma, Italy.

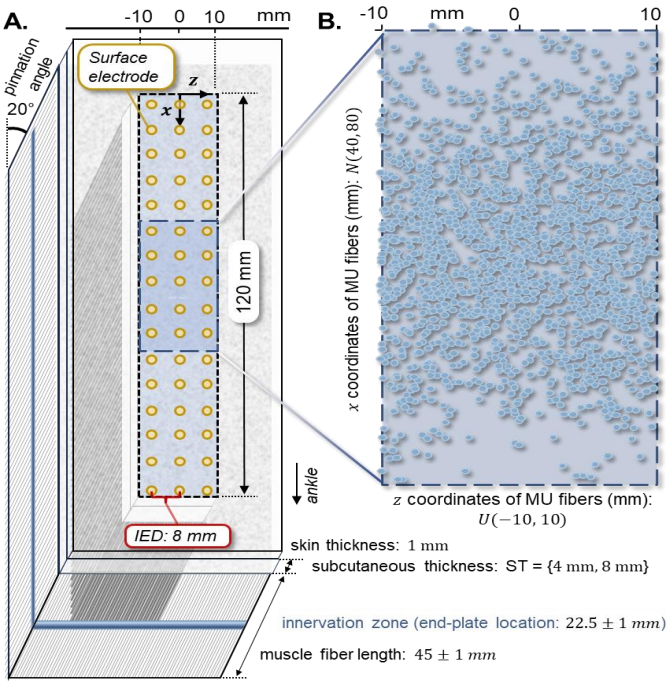

Fig 1: Simulated muscle volume and motor unit structure. Schematic representation of the simulated GM volume. The figure illustrates the simulation parameters and the spatial organization of fibers within a representative motor unit territory. Fibers are distributed along the longitudinal (Gaussian distribution with $\sigma = 10$ mm) and transverse (uniform distribution) axes, with Gaussian variability in end-plate location, fiber length, and conduction velocity.

alternation artifacts when MUs with similar thresholds respond variably to identical stimuli [2]. High-density surface electromyograms (HD-sEMG) has emerged as a potentially valid alternative to contend with these issues. This technique samples action potentials from MUs which fibers reside in different muscle regions [12], [13]. Such multiple detection has the potential to resolve alternation artifacts and to provide more representative sampling of muscle activity [14].

In this manuscript we therefore propose a novel MUNE methodology that synergizes HD-sEMG with artificial intelligence (AI). The present approach is compared with state-of-the-art methods, and is based on an extensive dataset of HD-sEMG simulated for medial gastrocnemius (MG): a key muscle for locomotion and postural stability [15]. We used these signals to train and validate a specialized dual-branch neural network (NN). The NN processes both raw HD-sEMG and their incremental changes through two complementary pathways (convolutional neural network (CNN) for spatial feature extraction and long short-term memory (LSTM) for temporal response modeling) providing stepwise predictions across the full stimulation range rather than a single global MUNE value. This allows for detailed insights into MU recruitment, spatial distribution, and neuromuscular organization. The proposed novel approach was tested using both simulated data and experimental signals from healthy subjects. As discussed here, our method represents a significant step toward more accurate and practical MUNE in both research and clinical fields.

II. Materials and Methods

First, we simulated HD-sEMG with a model of in-depth pinnate muscles that accurately reflects the anatomical and physiological MG properties [16]. The dataset was designed to reflect key biological and experimental sources of variability, including MU population size (nMUs), subcutaneous thickness (ST), and signal-to-noise ratio (SNR). Second, we designed the NN architecture combining convolutional and recurrent layers to predict the number of elicited MUs from the simulated signals. Finally, we compared HD-MUNet with two state-of-the-art MUNE methods on both simulated and experimental signals: Incremental MUNE [6] and StairFit MUNE [5]. The experimental procedures adhered to the Declaration of Helsinki and were approved by the local ethics committee (CER-Polito, Prot. No. 3152/2025). Full details on the simulated HD-sEMG and experimental protocol are described in the Appendix.

A. Simulated Data

1) Motor Unit Action Potentials

Surface EMG signals were simulated using the model developed by Mesin et al. [16]. The MG muscle was selected for its clinical relevance in posture and locomotion, as well as its suitability for surface EMG recordings [17], [18], [19]. The simulated muscle volume measures 120 mm (longitudinal) and 20 mm (transverse), with depth determined by fiber length (45 ± 1 mm; [20]) and pennation angle (20°; [19], [21]). Although smaller than the full anatomical muscle [22], this volume adequately represents the typical HD-sEMG recording region , with end-plate locations modeled with 1 mm jitter (Fig 1).

2) Motor Unit Recruitment

For each simulated subject, we implemented the inverted size principle to simulate MU response during an electrical stimulation protocol. Larger MUs, with lower excitation thresholds, responded before smaller ones. This approach stems from the biophysics principle that axons with larger diameters exhibit higher electrical conductivity and consequently lower excitability thresholds in response to external currents [23]. This produces threshold separation among large MUs at low intensities, and increasing overlap among smaller, more numerous MUs at higher intensities. The response threshold (RT [1]) for the i-th motoneuron was adapted from the voluntary contraction model proposed by Fuglevand et al. [24]:

$$RT(i) = RR - \mathrm{e}^{ai} \tag{1}$$

where i is the motoneuron index (from the smallest to the largest out of the n simulated MUs), RR is the recruitment range (set to 85 for broad recruitment conditions [24]), and a is the distribution coefficient given by $\ln(RR)/n$. Then, for each MU, we implemented a sigmoidal probability function centered at the response threshold $RT(i)$ (Fig 2), consistent with established MUNE methodologies [24]. To avoid placing thresholds exactly at the first or last stimulation step, the interval [0, RR] was normalized to [0, 100%] of the maximal stimulation intensities. The recruitment threshold of the biggest MU ($RT(n)$) was set at 5% of maximal stimulation intensity, while for the smallest MU we set $RT(1)$ at 95%.

[1] We use response threshold instead of recruitment threshold excitation, as conceived by Fuglevand et al (1993), because we simulate the condition in which units fire in response to a simulated external stimulus rather than centrally-generated, voluntary command.

$$P^i(s) = \frac{1}{1 + e^{-(s - RT(i))}}, \qquad s = 0\%, \dots, 100\% \tag{2}$$

where s represents the normalized stimulation intensity, ranging from 0% to 100% of the maximal stimulation. The resolution of s depends on the selected protocol, with a smaller increment between successive values when a larger number of stimulation steps (nSteps) is used.

Then, the sEMG ($y_{r,c}(t)$) was obtained as the summation of the MUAPs recruited at each step, plus additive noise ($w(t)$). The recruitment of the i-th MU was governed by the binary variable $\rho^i(s)$, which was set to 1 (active MU) when its sigmoidal probability $P^i(s)$ exceeded the value of a uniform random variable generated for each step $\gamma^i(s) \sim U(0,1)$.

$$y_{r,c}(t) = \sum_{i=1}^{n} MUAP^i_{r,c}(t)\, \rho^i(s) + w(t) \tag{3}$$

$$\rho^i(s) = \begin{Bmatrix} 1, & P^i(s) > \gamma^i(s) \\ 0, & P^i(s) < \gamma^i(s) \end{Bmatrix}, \qquad \gamma^i(s) \sim U(0,1) \tag{4}$$

Finally, we implemented three distinct stimulation protocols by using different nSteps (50, 100, 200). This approach allowed us to simulate varying clinical protocols while maintaining consistent recruitment dynamics across all conditions

3) Simulated DataSet

The simulation process generated a dataset of 5040 different subjects by systematically combining: two MU libraries (4mm and 8mm ST); seven MUs population scenarios (20, 50, 100, 150, 200, 250 and 300 MUs); three SNR values (0, 8 and 16 dB); and three stimulation protocols (50, 100 and 200 steps).

For each combination, we created 40 different subjects by introducing physiological variability in the number of fibers per MU, the spatial distribution of MU territories, and conduction velocity profile (inter and intra MUs), according to physiological evidence acknowledged above. We split the dataset into Training set (70%), Validation set (15%) for hyperparameter tuning, and Test set (15%) for final evaluation. The Test set was restricted to four nMUs (20, 50, 150 and 300 MUs), while the 100, 200 and 250 MUs were used exclusively during training and validation to ensure richer representation of the MU population space during learning.

To evaluate interpolation capabilities on unseen nMUs, we created an additional test set with three held-out populations (175, 225, 275 MUs) using nSteps = 200, ST = 4 mm, and SNR (0, 8, 16 dB). For each of the 9 combinations, 6 independent subjects were simulated with physiological variability, yielding 54 interpolation test samples.

B. Reliability Testing

To assess the reliability of HD-MUNet under realistic measurement conditions, we conducted a test–retest analysis on simulated data covering two complementary scenarios. Repeatability was evaluated by re-simulating six test subjects (ST = 8 mm, nSteps = 200), preserving their MUs architecture while generating new stochastic recruitment sequences and additive noise realizations, yielding 72 test–retest pairs across all nMU and SNR combinations. Robustness to electrode repositioning was assessed by simulating ±5 mm grid shifts (distally and proximally) relative to the original recording position, each yielding an additional 72 test–retest pairs.

C. HD-MUNet

We designed HD-MUNet, a custom neural network for HD-sEMG-based MUNE. The model combines convolutional and recurrent neural networks in a unique dual-branch design that processes both the raw sEMG and their incremental changes across stimulation steps (Fig 3). This approach allows the system to capture both the absolute spatial signal characteristics and their temporal evolution during the stimulation protocol.

The first input branch (M-wave branch) receives the vector $\boldsymbol{y} = \left(\langle y_{1,c}(t)\rangle, \dots, \langle y_{r,c}(t)\rangle\right)^T$, with $\langle \cdot \rangle$ indicating the average of sEMG (i.e., $y(t)$) across the three columns ($c = \{1,2,3\}$) and for each of the 16 rows ($r = \{1,..,16\}$) simulated (Fig 2). The

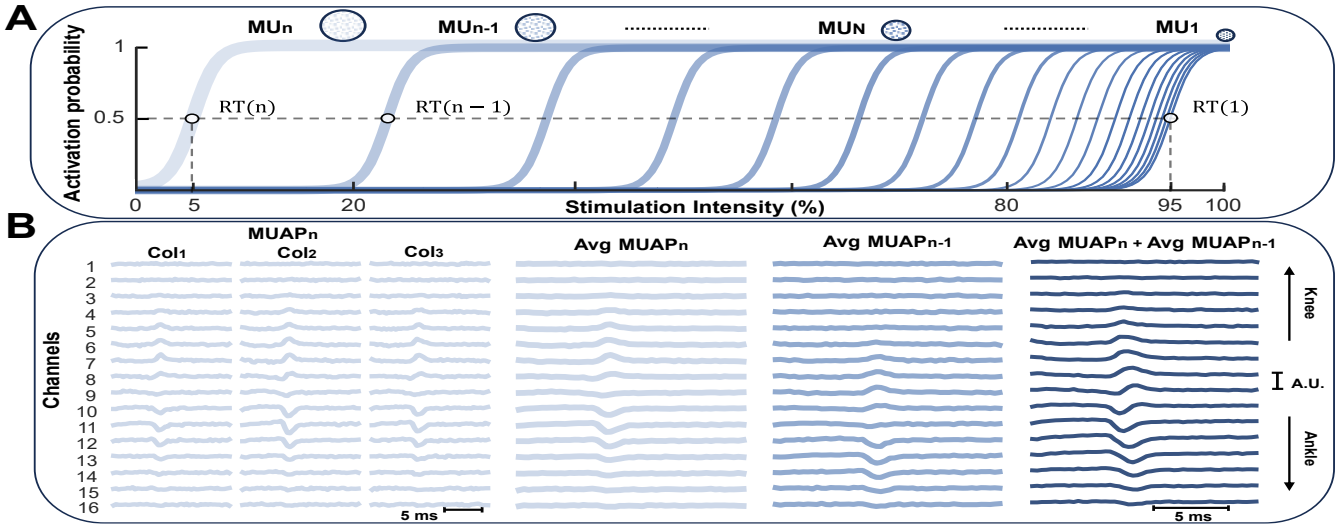

Fig 2: Motor unit recruitment and simulated HD-sEMG signals. (A) Recruitment probability curves of simulated MUs across stimulation steps. Larger MUs, with lower activation thresholds, are recruited at earlier and are widely separated in threshold. Conversely, smaller MUs, recruited at higher stimulation steps, exhibit closer thresholds. (B) Example of simulated HD-sEMG signals. From left to right: MUAP of the first recruited MU recorded with a 16×3 electrode matrix; the same MUAP averaged across electrode columns; MUAP of the second recruited MU averaged across columns; and the cumulative signal of the first two recruited MUs (column-averaged).

vector $\boldsymbol{y}$ can therefore be seen as an image, with 16 rows and the number of columns given by the sampling frequency and the window over which the response to stimulation, the M wave, is represented. In our simulations, these figures were respectively 4000 Hz and 12.5 ms (50 samples).

The Incremental branch analyzes step-to-step differences in muscle response, computed as:

$$y_k^{Inc} = y_k^{Mwave} - y_{k-1}^{Mwave}, \qquad k = \{2, \dots, nSteps\} \tag{5}$$

with the first step's incremental input set to zero and $nSteps$ denoting the number of stimulation steps. Both inputs have equal dimensions ($\boldsymbol{y} \in \mathbb{R}^{\mathrm{nSteps}\times 16\times 50}$). This symmetrical, dual-input strategy accounts for both constructive and destructive summation [25] of action potentials of previously and newly recruited MUs. Each branch passes through a CNN consisting of two convolutional layers with 3×3 kernels, ReLU activations, and max pooling operations, with global average pooling condensing spatial information into compact feature vectors. The extracted features from both branches are then concatenated and fed into a two-layer LSTM network (64 hidden units), which models temporal dependencies across the stimulation sequence. The final output provides continuous estimates of the cumulative number of elicited MUs at each stimulation step, with MUNE corresponding to the terminal step (Fig 3). Finally, to validate the importance of including both branches, we also trained two ablated architectures: a Raw-only model and an Incremental-only model. Both retained the CNN-LSTM structure but differed in input configuration.

HD-MUNet was trained with the Adam optimizer (in its PyTorch implementation), leveraging GPU acceleration and with the following hyperparameters: 500 epochs; $patience = 30\ epochs$; $learning\ rate = 10^{-4}$; $\varepsilon = 10^{-8}$; $\beta 1 = 0.9$; $\beta 2 = 0.99$; batch size $= 4\ subjects$; loss criterion $= L1\ Loss$. Computation was performed on a 13th Gen Intel(R) Core(TM) i7-13700H, 2400 MHz, 14 core.

D. Comparison with existing MUNE Methods

We tested two state-of-the-art solutions for MUNE using our dataset. For the "Incremental MUNE" method [6], representing the traditional MUNE approach, a single monopolar sEMG was generated out of the HD-sEMG. This single EMG was obtained by using a virtual electrode (10 mm of diameter) positioned at the center of the muscle volume. The signal simulated at the first step, with no MU elicited, was used for defining a minimal amplitude increase resulting from the excitation of MUs between two consecutive stimulation steps. After discarding increments below the noise threshold, the amplitude of the SMUP was estimated as the mean of the first 15 valid increments [2]. The MUNE value was obtained by dividing the maximal CMAP amplitude by this SMUP.

The recently proposed MUNE method we considered was the StairFit algorithm [5]. This method requires a $2 \times N$ input matrix $\boldsymbol{z}$, with N corresponding to the number of stimulation steps. The first and second rows of $\boldsymbol{z}$ are given by stimulation intensities, expressed in mA, and the corresponding response, expressed in mV, as provided by the large virtual electrode just described. The generated signal was then rescaled to fit the physiological MG ranges: 10-130 mA for stimulation and 0-11 mV for response. In addition, StairFit requires an initial value of the expected MUNE value (D1), which we set adaptively based on the number of simulated MUs: D1 = $0.8nMUs$ when $nMUs \leq nSteps$ and D1 = $nSteps - 10$ otherwise. The method models the CMAP scan as an ideal monotonically increasing staircase function, with each "stair" representing the cumulative amplitude of elicited action potential. The primary goal of the method is to find the optimal set stair heights that best fit the recorded CMAP amplitudes by minimizing the distance between each recorded point and its nearest "stair". The MUNE is then determined as the minimum number of MUs that can satisfy a predefined error requirement. All input parameters were kept at the default values recommended by Chen et al. [5].

E. Statistical Analysis

We evaluated the performance of each method separately (RStudio® version 4.4.2.). A four-way interaction linear mixed model ("lmer" RStudio® function) was applied to assess the effect of nMUs, nSteps, ST, and SNR on the percentage error relative to the simulated MU count for each subject. Positive and negative relative error values respectively indicate underestimation and overestimation. The hypothesis of normality of residuals was tested through the Shapiro-Wilk test ("shapiro" RStudio® function), while the hypothesis of homoscedasticity was tested through the Levene test of variance ("car::leveneTest" RStudio® function). If the normality and homoscedasticity hypothesis were satisfied, analysis of variance (ANOVA) with four-way repeated

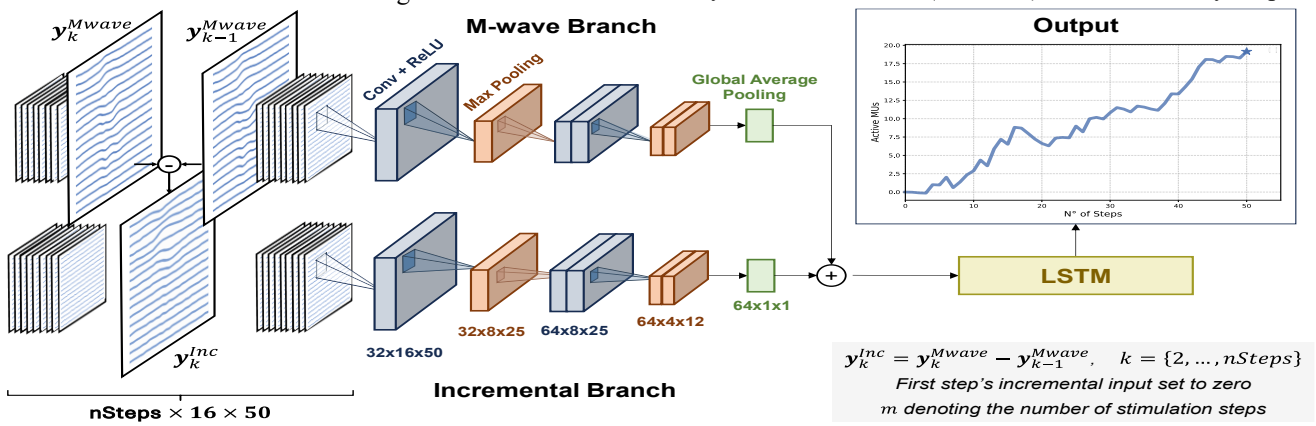

Fig 3: Architecture of the proposed HD-MUNet. The dual-branch network combines CNNs for spatial feature extraction and LSTM layers for temporal modeling. The two branches process raw HD-sEMG responses, and their outputs are merged to generate the final prediction. The network provides both the total MUNE value and the cumulative number of recruited MUs at each stimulation step.

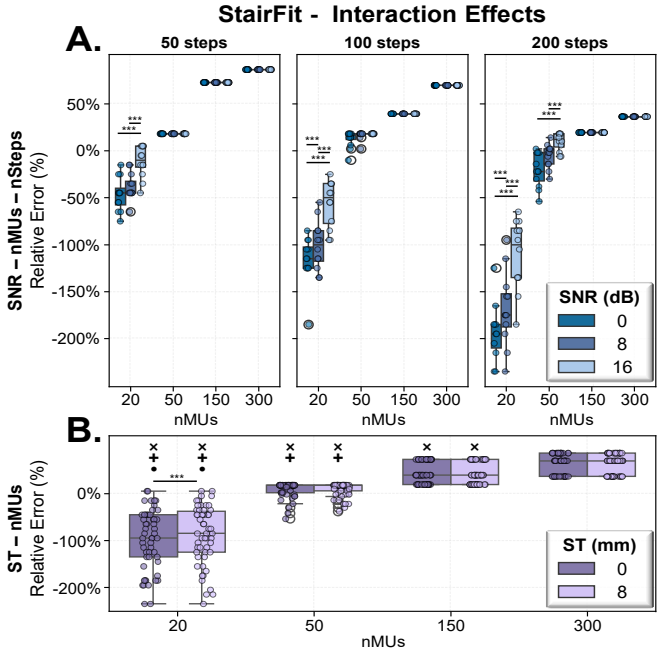

Fig 4: Interaction effects for StairFit. (A) SNR × nMUs × nSteps: with median relative errors ranging from −2.0% to −195%;(B) ST × nMUs: with large overestimations at 20 MUs, minimal errors at 50 MUs, and increasing overestimation for larger nMUs where estimates reflected only the value D1. ***Legend:*** *• significantly different from 50 MUs, + from 150 MUs, × from 300 MUs; *** p<0.0001.*

measures (nMUs, nSteps, ST and SNR) was performed to identify significant effects of the four factors and their interactions. Post hoc, pairwise comparisons were performed using the Bonferroni correction. After Q–Q plots visual inspection to assess residual normality, the experimental data set was analysed by following an identical procedural approach; however, the sole nSteps factor was considered. For test–retest scenarios, we computed Pearson correlation coefficients (r), coefficient of determination (R^2), and linear regression parameters (slope and intercept) to quantify agreement between initial and repeat estimates.

III. Results

A. Simulation Results

1) Main and Interaction Effects for StairFit

As shown in Fig 4, StairFit's performance fluctuated across all tested conditions. Different combinations of SNR, nMUs and nSteps resulted in markedly different errors (Fig 4A). Median relative errors ranged from -2.0% (8 dB × 50 MUs × 200 steps) to -195% (0 dB × 20 MUs × 200 steps). Overall, the 50 MUs × 200 steps condition emerged as the most stable, yielding relative errors of at most -22% across all SNR conditions. A clear dependency on the nSteps/nMUs ratio emerged: when this ratio was small (<1.5), the algorithm produced markedly positive errors due to the influence of the initialization parameter D1. Conversely, very large ratios (>5) consistently yielded strong negative errors, with overestimations exceeding 50% across all noise conditions.

Analysis of the significant nMUs × ST interaction (Fig 4B) revealed distinct patterns across MU populations. At 20 MUs, StairFit largely overestimated MUNE, by 95% at 4 mm and by 85% at 8 mm. In contrast, the 50 MUs condition showed the best performance, yielding only slight overestimations. For larger populations (150 and 300 MUs), estimates reflected only the initialization value D1.

2) Main and Interaction Effects for Incremental MUNE

Incremental MUNE exhibited three different significant interactions (Fig 5), indicating that estimation accuracy varied substantially across experimental conditions. Relative errors depended strongly on the combination of nMUs and nSteps (Fig 5A). Underestimation was pervasive across almost all conditions, with the only notable exception at 50 MUs × 200 steps (median error 20%). The 50 steps condition consistently produced the worst results, with errors exceeding 66% regardless of nMUs. Increasing the number of steps improved accuracy but introduced broader variability, particularly at low MU numbers. A similar dependency was observed between MU population and noise level (Fig 5B). The 0 dB condition consistently yielded the largest errors (above 65%), whereas 16 dB produced the most accurate but also the most variable outcomes. The most accurate configuration overall was 150 MUs × 16 dB, where the median error was 0%. The interaction between SNR and number of steps (Fig 5C) confirmed the strong dependence on both noise level and stimulation resolution. At 50 steps, the three SNR levels yielded comparable results, but at 100 and 200 steps differences across SNR conditions became pronounced, with the most accurate outcomes at 8 dB × 200 steps (9%) and 16 dB × 200 steps (5%).

3) Main and Interaction Effects for HD-MUNet

As shown in Fig 6, HD-MUNet exhibited only one significant interaction effect between nMUs and ST, with no significant effects observed for nSteps and SNR. Overall, errors remained below 7% (median values) across all tested conditions. At low nMUs counts, performance was remarkably

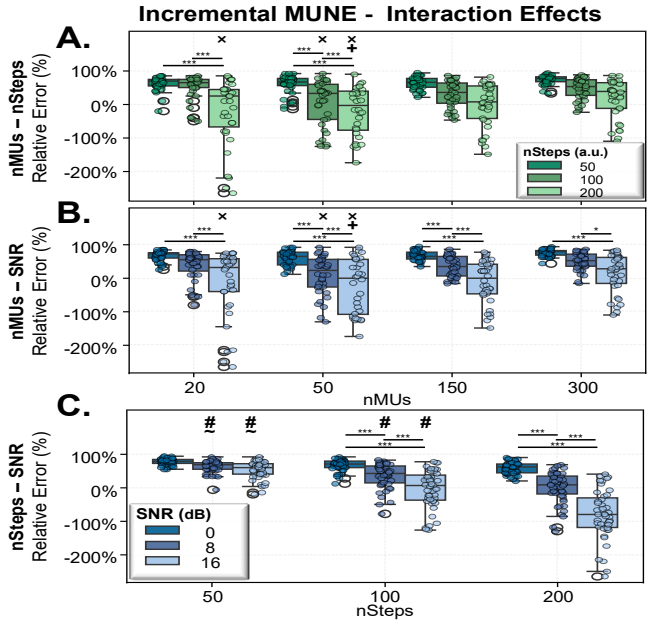

Fig 5: Interaction effects for Incremental MUNE. (A) nSteps × nMUs: underestimation across conditions; (B) SNR × nMUs: largest errors at 0 dB (>65%) and improved but more variable performance at 16 dB; (C) SNR × nSteps: showing comparable performance across SNRs at 50 steps and increasing divergence at 100 and 200 steps. ***Legend:*** *+ significantly different from 150 MUs, × from 300 MUs; ~ from 100 steps; # from 200 steps; ** p<0.05; *** p<0.0001.*

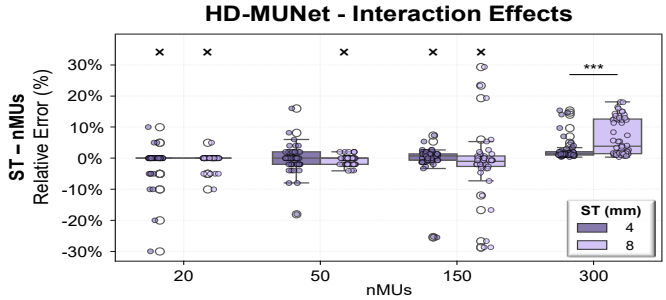

Fig 6: Interaction effects for HD-MUNet. ST × nMUs: the only significant interaction Increasing nMUs leads to increased variability, particularly in the 8 mm condition. Performance remains robust at low nMU counts with minimal ST sensitivity. At higher nMU counts, ST influences performance at 300 MUs. ***Legend:*** *× significantly different from 300 MUs; *** p<0.0001.*

similar between the two ST conditions (at 50 MUs median errors were -2.0% and -0.04% for 4 mm and 8 mm respectively), demonstrating robustness to ST variations. In contrast, ST influenced performance at higher MU counts, with significant differences detected exclusively at 300 MUs, where median errors were 2.5% (4 mm) and 6.7% (8 mm), accompanied by a progressive increase in variability (sd = 0.034 and 0.059 respectively). Increasing nMUs led to a progressive increase in variability, with a particularly pronounced underestimation in the 8 mm condition.

4) Ablation Study Analysis

As illustrated in Fig 7 The full two-branch model achieved best validation performance (MAE = 6.07 MU), with the lowest variability across training epochs. The Raw-only model performed substantially worse (MAE = 24.03 MU, +296%), demonstrating that M-wave information alone is insufficient for reliable MUNE estimation. While the Incremental-only model achieved an intermediate performance (MAE = 11.78 MU, +94%), indicating that step-to-step differences carry discriminative information but are insufficient when considered alone.

5) Reliability Testing

HD-MUNet demonstrated exceptional reliability under simulated test–retest conditions. Across 72 conditions generated by re-simulating six test subjects, the model achieved a Pearson correlation coefficient of r = 0.995. The regression slope was 1.02 [95% CI: 1.00, 1.05] and the intercept was -0.94 [95% CI: -4.78, 2.89] MU, indicating negligible systematic bias between test and retest estimates (Fig 8A).

Robustness to electrode repositioning was confirmed through simulated ±5 mm grid displacements (Fig 8B-C). For the distal shift (+5 mm), HD-MUNet maintained high accuracy with r = 0.997, slope = 0.97 [95% CI: 0.96, 0.99], and intercept = 0.44 [95% CI: -2.36, 3.24] MU. For the proximal shift (-5 mm), results were comparable: r = 0.995, slope = 1.04 [95% CI: 1.01, 1.06], and intercept = -0.96 [95% CI: -4.95, 3.03] MU.

6) Interpolation Capabilities

As shown in Fig 9 HD-MUNet demonstrated robust generalization. When tested on MU populations (175, 225, 275 MUs) not present during training, HD-MUNet achieved consistently low errors < 7%. In contrast, StairFit consistently returned D1 across all interpolated nMUs, yielding constant errors of (19.4, 30.6%) regardless of the actual MU count. Incremental MUNE exhibited high variability across interpolated conditions (4.22, 20.32% with std > 49%), particularly sensitive to SNR variations.

B. Experimental Results

For the experimental MUNE, the sole factor included in the statistical model was nSteps. Also, since the true number of MUs is unknown in vivo, absolute MUNE counts are reported. As shown in Fig 10, although all three methods showed a qualitative tendency toward larger MUNE when nSteps increased from 50 to 200, this trend reached statistical significance only for the Incremental and StairFit methods.

For HD-MUNet, no significant effect was detected for nSteps. Predictions remained stable across the 50, 100, and 200 steps conditions, indicating that the method is largely insensitive to variations in the stimulation protocol. In contrast, StairFit exhibited significant differences between 50 and 100 steps and between 50 and 200 steps, consistently returning the initialization value D1 at low step counts and diverging only at 200 steps. For the Incremental method, a significant difference emerged exclusively between the 50 and 200 steps conditions.

IV. DISCUSSION

This study introduces HD-MUNet, a novel MUNE methodology that integrates HD-sEMG with AI. The novelty of the method consists in a cutting edge NN architecture that combines CNNs and LSTMs in a dual-branch design that processes both the raw M waves and their incremental changes across stimulation steps. The rationale behind this design was to overcome key limitations of existing approaches, including spatial under sampling of MU activity [26], and the vulnerability to alternation artefacts [1]. Simulation results showed that HD-MUNet achieved consistently low errors across all tested conditions, with performance independent from SNR and stimulation protocol. In contrast, state-of-the-art methods showed substantially greater and more variable errors, with strong sensitivity to SNR, nMUs, and nSteps

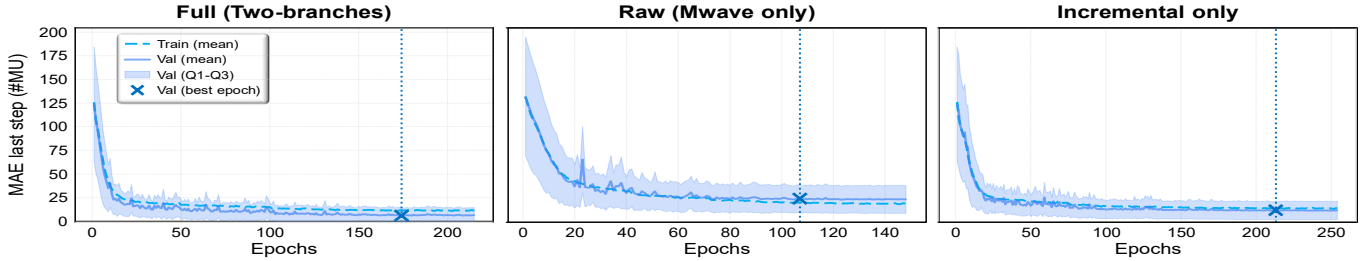

Fig 7: Ablation Study: Contribution of M-wave and Incremental Branches. Full two-branch model showing optimal convergence and low validation error; Raw (M-wave only) model: significant performance degradation (+296% MAE); Incremental only model: substantial performance loss (+94% MAE). Learning curves demonstrate that both branches are essential for accurate MUNE estimation.

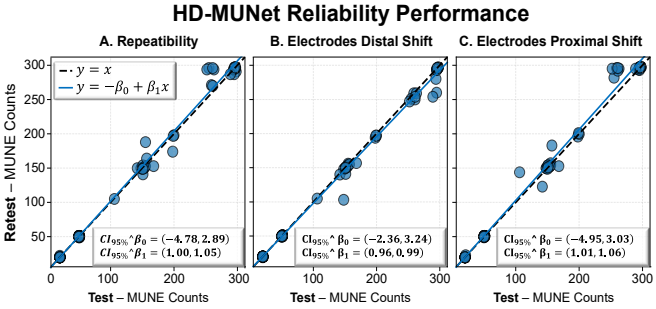

Fig 8: HD-MUNet Reliability Performance. HD-MUNet demonstrates exceptional repeatability under simulated test–retest conditions: stochastic recruitment (r = 0.9955), distal electrode shift (r = 0.9973), and proximal electrode shift (r = 0.9952). Strong linearity with slopes (~1.0) and minimal intercepts confirms stability.

A. Motivation for the two state-of-the-art MUNE tested

Among the many MUNE techniques available, we focused our comparison on Incremental MUNE and StairFit MUNE [5], [6]. Incremental MUNE was included as the most common approach, providing a natural baseline for comparison. Conversely, StairFit represented the most recent MUNE method at the time of our study, with an implementation that was openly available. Alternative MUNE methods, such as HD-sEMG based approaches, were not considered because of their assumption that the mean amplitude of single MU action potential is comparable across the MU population. However, this assumption is not fully consistent with physiological parameters [27]. In a healthy muscle, and more so under pathological conditions involving reinnervation, MU sizes can vary substantially [28].

B. Single detection points yield biased and highly variable MUNE values

The Incremental and StairFit methods, based on a single detection point, demonstrated competitive performance only under favourable simulated conditions.

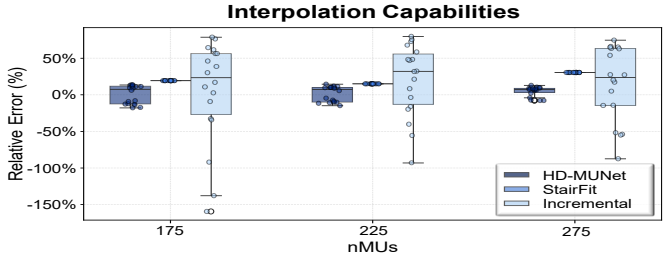

Fig 9: HD-MUNet Interpolation capabilities. HD-MUNet achieves ≤ 5% error on interpolated nMUs population (175, 225, 275 MUs), while StairFit returns constant values (30.6%) and Incremental MUNE shows high variability (std > 49%).

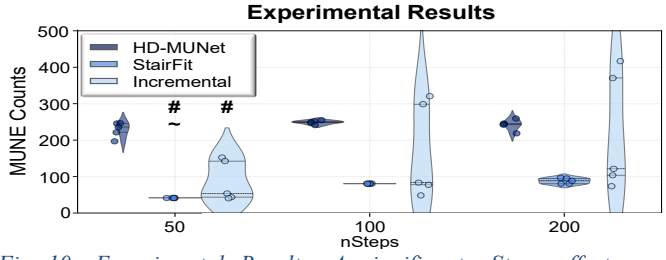

Fig 10: Experimental Results. A significant nSteps effect was observed for Incremental and StairFit, but not for HD-MUNet. ***Legend:*** *~ significantly different from 100 steps; # from 200 steps.*

The analysis of StairFit highlighted structural limitations that are closely related to its reliance on CMAP scan data [5]. When nSteps was insufficient relative to the number of MUs simulated, the algorithm was unable to produce a true MUNE estimate and systematically returned the D_1. This consistent behavior was expected, as CMAP-scan based methods, such as StairFit, typically require a substantially larger number of stimulation steps (500) to reliably resolve MU recruitment [1], [11]. Within operable conditions (nSteps/nMUs ratio between 2 and 4, high SNR), errors were comparable to HD-MUNet (within ±15%), though with greater variability and sensitivity to signal quality and protocol configuration. At low SNR, StairFit consistently overestimated MUNE, likely due to the misclassification of noise fluctuations as recruitment events. Similarly, when the nSteps/nMUs ratio became excessively large, pronounced overestimation emerged, consistent with alternation phenomena whereby the same MU is repeatedly recruited and derecruited across fine stimulation steps and erroneously counted as multiple units [2]. An additional limitation emerged with respect to ST. At 20 MUs, the only nMUs for which StairFit produced estimates across all step conditions, a significant difference was observed between the two ST values. This finding suggests sensitivity to tissue filtering effects.

On the other side, the Incremental MUNE method exhibited a strong dependence on nMUs, nSteps, and SNR, while showing no detectable sensitivity to ST. This latter finding is consistent with its reliance on relative CMAP increments rather than absolute signal amplitude [6]. Across most simulation conditions Incremental MUNE tended to underestimate MUNE (Fig 5), particularly at low nSteps (50 steps) and low SNR (0dB). This behavior is intrinsic to the method's formulation, as MUNE is derived from the ratio between maximal CMAP amplitude and the mean SMUP [6]. With few steps or poor signal quality, detected increments tend to be large, overestimating mean SMUP and consequently underestimating MUNE. Conversely, at high nSteps and SNR, very fine increments unrelated to true recruitment creates a bias in unit count, leading to overestimation, an effect most evident at low nMUs where alternation is more likely. This effect was most evident in conditions combining high stimulation resolution (200 steps) and low nMUs (20 MUs) where alternation phenomena are more likely to occur [2]. Overall, these findings indicate that Incremental MUNE is highly sensitive to experimental configuration and signal quality. While increasing the number of steps and improving SNR can reduce systematic underestimation, these same conditions may introduce substantial variability and severe overestimation due to alternation effects, particularly in small MU populations [1].

Taken together, these results indicate that StairFit and Incremental MUNE can provide reasonable estimates only under a restricted set of favorable conditions. These dependencies limit the robustness and generalizability of the methods.

C. HD-MUNet provides robust and accurate MUNE values

We believe the main determinant of the highly variable performance of the two state-of-the-art methods we considered was due to their use of a single detection point. Recently, high-density surface EMG has been shown to enable spatially

distributed analysis of motor unit activity from multiple recording sites over the skin surface [26]. This is particularly true for in-depth pinnate muscles [29], like gastrocnemius. It is therefore conceivable that MUs which fibers reside far away from the detection point does not contribute to the detected EMGs and therefore to MUNE. The proposed HD-MUNet architecture was specifically designed to exploit HD-sEMG spatiotemporal information. CNN extracts spatial activation patterns across the electrode grid, thereby capturing features in MUs territories that would be missed in single-channel acquisitions. In parallel, LSTM models temporal dependencies across successive stimulation steps, enabling the network to learn recruitment dynamics. The dual-branch design, translated into highly stable performance across all simulated conditions, with consistently low relative errors and a predominant tendency toward mild underestimation. Test-retest analysis confirmed that estimates were highly reproducible under repeated simulations (Fig 8). HD-MUNet robustness further extended to electrode repositioning (±5 mm). Critically, errors on held-out nMUs were comparable to performance on trained populations, indicating that the network learned generalizable features rather than memorizing training data.

Notably, HD-MUNet was not affected by nSteps and SNR, indicating that stimulation protocol and noise level do not influence its predictions. However, a significant nMUs × ST interaction emerged, revealing that ST affects performance at high nMUs. Performance was largely stable up to 150 MUs, with median errors within ±3% across all ST conditions. At 300 MUs, a tendency toward positive errors, reflecting mild underestimation, was observed across both ST values. This is likely attributable to the limited nSteps relative to nMUs, which makes it increasingly difficult to resolve individual MUs. Overestimation was more pronounced at 8 mm ST, presumably due to a reduced discriminability of individual MUs due to the ST low-pass filter effect [30], [31].

It should be noted that HD-MUNet's ability to distinguish MU populations when nSteps < nMUs relies on morphological differences in CMAP characteristics shaped by physiologically recruitment threshold distributions. Nevertheless, even in this worst-case scenario, median relative errors remained below 7%, indicating that HD-MUNet preserves reliable estimation accuracy across the full range of simulated conditions.

Beyond the global MUNE estimate, HD-MUNet outputs a stepwise recruitment curve that maps the cumulative number of elicited MUs across the full stimulation range. This representation may provide clinically relevant information beyond a single MUNE. For instance, the stimulation intensity at which a given fraction of the MU pool is recruited (an analog of the D50 parameter derived from CMAP scans [7]) could serve as a sensitive marker of neuromuscular organization. In neurodegenerative conditions such as ALS, and following neurological injury such as stroke, the shape of the stimulus-response curve is known to change with MU loss and reinnervation [1], [3], [32]. A recruitment curve expressed directly in MU counts, rather than CMAP amplitude, may therefore offer complementary information for monitoring disease progression, though this application requires dedicated validation in patient populations.

D. Performance on Experimental Data

Experimental data confirmed the simulation findings. StairFit returned the initialization value D_1 at both 50 and 100 steps, diverging only at 200 steps. Incremental MUNE showed a significant dependence on nSteps between the 50 and 200 steps conditions. HD-MUNet predictions remained stable across all step counts, confirming its robustness to stimulation protocol variations in vivo. Beyond sensitivity to stimulation parameters, the interpretation of absolute MUNE values deserves careful consideration. By definition, only MUs whose fibers lie within the pick-up volume of the electrode grid contribute to MUNE, and the sampled muscle volume inevitably varies across subjects due to anatomical differences. At 200 steps, three subjects with similar muscle lengths (21-23 cm) showed comparable HD-MUNet estimates (243-246 MUs), consistent with a similar proportion of the MU pool being sampled by the electrode grid. Subject 3 (25.5 cm) exhibited lower estimates (219), in line with a longer muscle reducing the relative coverage of the grid. However, Subject 2 (25 cm), with a comparable muscle length, showed the highest estimates across all subjects (260). Inter-subject variability in absolute MUNE values likely reflects a combination of anatomical factors (e.g. ST, muscle cross-sectional area, pennation angle) and protocol-related factors, including achieving true supramaximal stimulation for each subject.

V. Conclusion

In conclusion, HD-MUNet exhibited consistent performance characterized by minimal variability in relative error across a broad spectrum of simulated conditions, contrasting sharply with the instability observed in traditional approaches. High test-retest reliability and robustness to electrode repositioning further support its suitability for repeated clinical assessments. In addition to providing a global MUNE estimate at maximal stimulation, HD-MUNet outputs a stepwise recruitment curve across the entire stimulation range, with potential to extend the scope of MUNE beyond a single index of muscle activity toward a more detailed characterization of MU organization and recruitment dynamics. Nevertheless, the present study has two main limitations. First, HD-MUNet was validated using a single muscle geometry. Second, simulations assumed physiologically regular recruitment threshold distributions based on the Henneman size principle. Further validation across different muscle architectures and pathological recruitment patterns, such as those observed in ALS or myopathies, will be required to fully assess its robustness in clinical applications.

Appendix

Supplementary material associated with this article is available online.

Acknowledgment

The authors would like to acknowledge the support received within the «BREAKing BONDS» project – funded by the Ministero dell'Università e della Ricerca under the PRIN 2022 program (D.D. 104 - 02/02/2022). This manuscript reflects only the authors' views and opinions, and the Ministry cannot be held responsible for them.

REFERENCES

[1] C. L. Gooch *et al.*, «Motor unit number estimation: A technology and literature review», *Muscle Nerve*, vol. 50, fasc. 6, pp. 884–893, 2014, doi: 10.1002/mus.24442.

[2] M. De Carvalho, P. E. Barkhaus, S. D. Nandedkar, e M. Swash, «Motor unit number estimation (MUNE): Where are we now?», *Clin. Neurophysiol.*, vol. 129, fasc. 8, pp. 1507–1516, ago. 2018, doi: 10.1016/j.clinph.2018.04.748.

[3] C. Neuwirth *et al.*, «Motor Unit Number Index (MUNIX) detects motor neuron loss in pre-symptomatic muscles in Amyotrophic Lateral Sclerosis», *Clin. Neurophysiol. Off. J. Int. Fed. Clin. Neurophysiol.*, vol. 128, fasc. 3, pp. 495–500, mar. 2017, doi: 10.1016/j.clinph.2016.11.026.

[4] K. J. Swoboda *et al.*, «Natural history of denervation in SMA: relation to age, SMN2 copy number, and function», *Ann. Neurol.*, vol. 57, fasc. 5, pp. 704–712, mag. 2005, doi: 10.1002/ana.20473.

[5] M. Chen, Z. Lu, Y. Zong, X. Li, e P. Zhou, «A Novel Analysis of Compound Muscle Action Potential Scan: Staircase Function Fitting and StairFit Motor Unit Number Estimation», *IEEE J. Biomed. Health Inform.*, vol. 27, fasc. 3, pp. 1579–1587, mar. 2023, doi: 10.1109/JBHI.2022.3229211.

[6] A. J. McComas, P. R. Fawcett, M. J. Campbell, e R. E. Sica, «Electrophysiological estimation of the number of motor units within a human muscle», *J. Neurol. Neurosurg. Psychiatry*, vol. 34, fasc. 2, pp. 121–131, apr. 1971, doi: 10.1136/jnnp.34.2.121.

[7] J. H. Blok, A. Ruitenberg, E. M. Maathuis, e G. H. Visser, «The electrophysiological muscle scan», *Muscle Nerve*, vol. 36, fasc. 4, pp. 436–446, 2007, doi: 10.1002/mus.20838.

[8] T. J. Doherty, D. W. Stashuk, e W. F. Brown, «Multiple point stimulation and F-response MUNE techniques», in *Supplements to Clinical Neurophysiology*, vol. 55, in Motor Unit Number Estimation (MUNE): Proceedings of the First International Symposium on MUNE, vol. 55. , Elsevier, 2003, pp. 29–40. doi: 10.1016/S1567-424X(02)55004-2.

[9] M. B. Bromberg, «Spike triggered averaging MUNE technique», in *Supplements to Clinical Neurophysiology*, vol. 55, in Motor Unit Number Estimation (MUNE): Proceedings of the First International Symposium on MUNE, vol. 55. , Elsevier, 2003, pp. 97–107. doi: 10.1016/S1567-424X(02)55011-X.

[10] P. G. Ridall, A. N. Pettitt, R. D. Henderson, e P. A. McCombe, «Motor Unit Number Estimation—A Bayesian Approach», *Biometrics*, vol. 62, fasc. 4, pp. 1235–1250, 2006, doi: 10.1111/j.1541-0420.2006.00577.x.

[11] H. Bostock, «Estimating motor unit numbers from a CMAP scan», *Muscle Nerve*, vol. 53, fasc. 6, pp. 889–896, giu. 2016, doi: 10.1002/mus.24945.

[12] D. Farina e A. Holobar, «Characterization of Human Motor Units From Surface EMG Decomposition», *Proc. IEEE*, vol. 104, fasc. 2, pp. 353–373, feb. 2016, doi: 10.1109/JPROC.2015.2498665.

[13] T. M. Vieira e A. Botter, «The Accurate Assessment of Muscle Excitation Requires the Detection of Multiple Surface Electromyograms», *Exerc. Sport Sci. Rev.*, vol. 49, fasc. 1, p. 23, gen. 2021, doi: 10.1249/JES.0000000000000240.

[14] J. P. Van Dijk, J. H. Blok, B. G. Lapatki, I. N. Van Schaik, M. J. Zwarts, e D. F. Stegeman, «Motor unit number estimation using high-density surface electromyography», *Clin. Neurophysiol.*, vol. 119, fasc. 1, pp. 33–42, gen. 2008, doi: 10.1016/j.clinph.2007.09.133.

[15] I. D. Loram, C. N. Maganaris, e M. Lakie, «Human postural sway results from frequent, ballistic bias impulses by soleus and gastrocnemius», *J. Physiol.*, vol. 564, fasc. Pt 1, pp. 295–311, apr. 2005, doi: 10.1113/jphysiol.2004.076307.

[16] L. Mesin, R. Merletti, e T. M. M. Vieira, «Insights gained into the interpretation of surface electromyograms from the gastrocnemius muscles: A simulation study», *J. Biomech.*, vol. 44, fasc. 6, pp. 1096–1103, apr. 2011, doi: 10.1016/J.JBIOMECH.2011.01.031.

[17] M. V. Narici, T. Binzoni, E. Hiltbrand, J. Fasel, F. Terrier, e P. Cerretelli, «In vivo human gastrocnemius architecture with changing joint angle at rest and during graded isometric contraction.», *J. Physiol.*, vol. 496, fasc. 1, pp. 287–297, 1996, doi: 10.1113/jphysiol.1996.sp021685.

[18] I. Di Giulio, C. N. Maganaris, V. Baltzopoulos, e I. D. Loram, «The proprioceptive and agonist roles of gastrocnemius, soleus and tibialis anterior muscles in maintaining human upright posture», *J. Physiol.*, vol. 587, fasc. Pt 10, pp. 2399–2416, mag. 2009, doi: 10.1113/jphysiol.2009.168690.

[19] T. M. M. Vieira, I. D. Loram, S. Muceli, R. Merletti, e D. Farina, «Postural activation of the human medial gastrocnemius muscle: Are the muscle units spatially localised?», *J. Physiol.*, vol. 589, fasc. 2, pp. 431–443, 2011, doi: 10.1113/JPHYSIOL.2010.201806.

[20] R. S. Chow, M. K. Medri, D. C. Martin, R. N. Leekam, A. M. Agur, e N. H. McKee, «Sonographic studies of human soleus and gastrocnemius muscle architecture: gender variability», *Eur. J. Appl. Physiol.*, vol. 82, fasc. 3, pp. 236–244, 2000, doi: 10.1007/S004210050677.

[21] P. A. Huijing, «Architecture of the Human Gastrocnemius Muscle and Some Functional Consequences», *Acta Anat. (Basel)*, vol. 123, fasc. 2, pp. 101–107, lug. 2008, doi: 10.1159/000146047.

[22] K. Andjelkov, T. C. Atanasijevic, V. M. Popovic, M. Sforza, C. J. Atkinson, e I. Soldatovic, «Anatomical aspects of the gastrocnemius muscles: A study in 47 fresh cadavers», *J. Plast. Reconstr. Aesthet. Surg.*, vol. 69, fasc. 8, pp. 1102–1108, ago. 2016, doi: 10.1016/j.bjps.2016.04.002.

[23] E. Henneman, «Relation between size of neurons and their susceptibility to discharge», *Science*, vol. 126, fasc. 3287, pp. 1345–1347, dic. 1957, doi: 10.1126/science.126.3287.1345.

[24] F. Fuglevand Andrew, DA Winter, e AE Patla. «Fuglevand, A. J. ,. Winter, D. A. &. Patla, A. E. Models of recruitment and rate coding organization in motor-unit pools. J. Neurophysiol. 70, 2470-2488». Journal of neurophysiology 70 (dicembre 1993): 2470–88. https://doi. org/10. 1152/jn. 1993. 70. 6. 2470., D. Winter, e A. Patla, «Fuglevand, A. J., Winter, D. A. & Patla, A. E. Models of recruitment and rate coding organization in motor-unit pools. J. Neurophysiol. 70, 2470-2488», *J. Neurophysiol.*, vol. 70, pp. 2470–88, dic. 1993, doi: 10.1152/jn.1993.70.6.2470.

[25] K. G. Keenan, D. Farina, K. S. Maluf, R. Merletti, e R. M. Enoka, «Influence of amplitude cancellation on the simulated surface electromyogram», *J. Appl. Physiol.*, vol. 98, fasc. 1, pp. 120–131, gen. 2005, doi: 10.1152/japplphysiol.00894.2004.

[26] R. Merletti, A. Holobar, e D. Farina, «Analysis of motor units with high-density surface electromyography», *J. Electromyogr. Kinesiol. Off. J. Int. Soc. Electrophysiol. Kinesiol.*, vol. 18, fasc. 6, pp. 879–890, dic. 2008, doi: 10.1016/j.jelekin.2008.09.002.

[27] D. Farina, C. Cescon, e R. Merletti, «Influence of anatomical, physical, and detection-system parameters on surface EMG», *Biol. Cybern.*, vol. 86, fasc. 6, pp. 445–456, giu. 2002, doi: 10.1007/s00422-002-0309-2.

[28] M. Lukács, L. Vécsei, e S. Beniczky, «Changes in muscle fiber density following a stroke», *Clin. Neurophysiol.*, vol. 120, fasc. 8, pp. 1539–1542, ago. 2009, doi: 10.1016/j.clinph.2009.06.001.

[29] T. Vieira, G. L. Cerone, A. Botter, K. Watanabe, e A. Vigotsky, «The Sensitivity of Bipolar Electromyograms to Muscle Excitation Scales With the Inter-Electrode Distance», *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. PP, pp. 1–1, ott. 2023, doi: 10.1109/TNSRE.2023.3325132.

[30] D. S. de Oliveira *et al.*, «Neural decoding from surface high-density EMG signals: influence of anatomy and synchronization on the number of identified motor units», *J. Neural Eng.*, vol. 19, fasc. 4, p. 046029, ago. 2022, doi: 10.1088/1741-2552/AC823D.

[31] C. Cescon, P. Rebecchi, e R. Merletti, «Effect of electrode array position and subcutaneous tissue thickness on conduction velocity estimation in upper trapezius muscle», *J. Electromyogr. Kinesiol. Off. J. Int. Soc. Electrophysiol. Kinesiol.*, vol. 18, fasc. 4, pp. 628–636, ago. 2008, doi: 10.1016/j.jelekin.2007.01.005.

[32] M. Chen *et al.*, «Compound muscle action potential (CMAP) scan examination of paretic and contralateral muscles reveals motor unit alterations after stroke», *Sci. China Life Sci.*, vol. 66, fasc. 11, pp. 2604–2613, nov. 2023, doi: 10.1007/s11427-022-2308-8.