XAI for myo-controlled prosthesis: Explaining EMG data for hand Gesture classification

Noemi Gozzi, Lorenzo Malandri,Fabio Mercorio,Alessandra Pedrocchi

Knowledge-BasedSystems240(2022)108053

https://doi.org/10.1016/j.knosys.2021.108053

XAI for myo-controlled prosthesis: explaining EMG data for hand gesture classification

Noemi Gozzi[a], Lorenzo Malandri[b,*], Fabio Mercorio[b], Alessandra Pedrocchi[a]

[a] *Dept of Electronics, Information and Bioengineering, Politecnico di Milano, Milan, Italy*
[b] *Dept of Statistics and Quantitative Methods, University of Milano-Bicocca, Milan, Italy*

Abstract

Machine Learning has recently found a fertile ground in EMG signal decoding for prosthesis control. However, its understanding and acceptance are strongly limited by the notion of AI models as black-boxes. In critical fields, such as medicine and neuroscience, understanding the neurophysiological phenomena underlying models' outcomes is as relevant as classification performance. In this work, we adapt state-of-the-art XAI algorithms to EMG hand gesture classification to understand the outcome of machine learning models with respect to physiological processes, evaluating the contribution of each input feature to the prediction and showing that AI models recognize the hand gestures by mapping and fusing efficiently high amplitude activity of synergic muscles.

This allows us to (i) drastically reduce the number of required electrodes without a significant loss in classification performances, ensuring the suitability of the system for a larger population of amputees and simplifying the realization of near real-time applications and (ii) perform an efficient selection of features based on their classification relevance, apprehended by the XAI algorithms. This feature selection leads to classification improvements in term of robustness and computational time, outperforming correlation based methods. Finally, (iii) comparing the physiological explanations produced by the XAI algorithms with the experimental setting highlights inconsistencies in the electrodes positioning

*Corresponding author
Email address: lorenzo.malandri@unimib.it (Lorenzo Malandri)

over different rounds or users, then improving the overall quality of the process.

Keywords: EMG signal decoding, eXplainable AI, myo-controlled prosthesis

1. Introduction and Motivation

Myo-controlled control for upper limb prostheses is commonly used in clinical prosthetic systems, and have proven to be effective in replacing missing functions providing conscious control of the movement [1, 2]. Myo-controlled prostheses are artificial devices intended to restore the normal functions of a missing limb, using the residual neuromuscular activity (EMG) as a control signal. EMG signal is the electrical activity of a muscle in response to a stimulus from the motor cortex in the central nervous system. The EMG electrical activity can be measured with non-invasive surface electrodes placed on the body surface and used to classify movement intention or production. Classical Machine Learning [3, 4, 5, 6, 7, 8] and Deep Learning [9, 6, 10, 11] algorithms have recently shown promising performances in surface EMG pattern recognition for hand movement classification. Nevertheless, the acceptance of AI in prosthesis control is strongly limited by the intrinsic black-box nature of AI models and the trade-off between performance and interpretability [12, 13]. In critical fields such as medicine and neuroscience, reaching a deep understanding of the neurophysiological phenomena underlying model outcomes is as relevant as the classification performances. EMG signal represents the output of a brain processing in the motor cortex; gaining insights about how it works and how synergic muscles perform complex movements is fundamental to eventually use the classified signals properly in the control of prosthetic devices. EXplainable AI (XAI) may be a powerful tool to interpret AI models for EMG pattern recognition with respect to physiological processes and phenomena under study and, to the best of our knowledge, it is still an unexplored field.

The benefits of XAI in EMG pattern recognition are not only confined to open the black-box itself. XAI can address different challenges, such as feature selection and reduction [4, 14, 3], curse of dimensionality, choice of the classifier

[5, 6, 9, 15, 10] and the number of channels [3, 5, 6, 16, 17], electrode placements and number of movements (which impact a lot the easiness of the experimental acquisitions) [3, 5] that are all open questions in the field.

In this research, we show that, with the aim of Explainable AI (XAI), it is possible to (i) reduce the number of classification features maintaining competitive classification performances and (ii) simplify the EMG electrodes' configuration. While machine learning methods have reached high accuracies in pattern recognition-based myoelectric control, their computational complexity limits their use in real-time applications [18, 19]. Reducing the number of features, and therefore the computational complexity of the classification, without a drop in performances lead to more robust and reliable models and and paves the way for new real-time applications. Furthermore, the number of EMG electrodes to be used in such applications is a pivotal parameter. While some researchers claim that using a large number of sensors improves the classification accuracy [20, 3], at the same time, it can be uncomfortable for the user [18] and, in some cases, it might even be impractical due to the limited usable surface of the amputee [3, 6]. Moreover, having several EMG sensors limit a large number of real-world applications because of computational complexity, practicality, and cost of the hardware [21, 22].

This paper aims to investigate the functioning of black-box models to classify hand movements and to simplify the system settings, mainly addressing the following questions:

Q1: Can we explain through XAI, and with physiologically plausible explanations, a black-box model for the classification of EMG hand movement?

Q2: Can we exploit XAI to simplify the electrodes' setting, reducing hardware and software complexity and improving the comfort and usability of the device?

Q3: Can XAI drive improvements towards real applications of myo-controlled prostheses for amputees?

1.1. Contribution

We propose to apply XAI to the field of EMG signal decoding for myo-controlled prosthesis, providing physiologically highly plausible local and global explanations of how AI models classify hand gestures. AI is combined with domain expertise in a *"human in the loop"* approach that is fundamental to effectively understand the explanations in the specific application domain. In particular, our contributions can be summarized as:

1. **We apply XAI to the EMG hand movement classification task** to ensure explainability and causability of the models, providing highly plausible explanations of how black-boxes classify EMG. We succeed in revealing that AI models recognize the hand gestures by mapping and fusing efficiently high amplitude activity of synergic muscles.
2. **We exploit XAI results to perform feature selection** reducing the number of features without a drop in classification metrics, leading to classification improvements in term of robustness and computational time and showing that this method outperforms correlation based methods.
3. **We reduce the complexity of electrodes' configuration**, reducing the hardware and software complexity while achieving competitive performances. This allows to (i) make the system suitable for a larger population of amputees and (ii) reduce computational times and complexity, thus simplifying the realization of near real-time applications.
4. **We use the XAI insights to detect inconsistencies in the placement of electrodes**, comparing the activated electrodes with domain knowledge.

2. Related works

In this section, we review the relevant literature related to the main topics paper: EMG pattern recognition and Explainable AI (XAI).

2.1. EMG pattern recognition

Recently, thanks to the availability of large amounts of data and high processing capabilities, data-driven approaches such as Machine Learning and Deep

Learning are achieving good performances for EMG signal decoding. Machine Learning models, such as LDA, SVM, and Tree Ensembles, require a feature selection step. Their performances rely on the choice of an optimal number and set of features and on the selection of a correct number and setting of electrodes. Phinyomark et al. [4, 14] presented a detailed study of all possible features for EMG decoding, highlighting how many research works have explored different feature vectors for EMG decoding but without evaluating their quantitative information and redundancy. In their first work [4], they recommended Mean Absolute Value (MAV), Waveform Length (WL), Wilson Amplitude (WAMP), AutoRegressive coefficients (AR), and Mean Absolute Value Slope (MAVS) as the most informative features derived from statistical analysis. However, in [14] a different set of four features, including Sample Entropy (SampEn), the fourth order cepstrum coefficients (CC), root mean square (RMS) and WL, achieved the best performance in the classification of ten upper limb motions.Furthermore, their approach for feature selection did not rely on class-specific feature contributions, and they did not discuss the number of electrodes. Al-Timemy et al. [3], starting from coefficients of 6^{th} order AR model, RMS, WL, Zero Crossing (ZC), and integral absolute value and Slope Sign Change (SSC), applied Principal Component Analysis (PCA) and Orthogonal Fuzzy Neighbourhood Discriminant Analysis (OFDNA) to the original set of features for feature reduction and classified using LDA and SVM models. In addition, they argued that limiting the number of electrodes is fundamental to increase the usability of sEMG systems because several amputees have limited body surface where to place electrodes. They removed the worst channel at each step by computing the accuracy without that channel. However, their empirical channel elimination technique may lead to a sub-optimal selection of electrodes. They proposed the ratio between the number of movements and electrodes as an index of the complexity of a system. Rehman et al. [6] proposed a classification with LDA and autoencoders achieving good performances with only four features: MAV, WL, SSC and ZC; however, features selection is not discussed with a reduction of sensors. Another major challenge is EMG variability, deeply

analyzed in [5, 6, 8]; EMG patterns vary depending on electrode shift, subject's mobility, changes in arm posture, and fatigue among other causes and these differences affect classification performances. The capability of generalization over patterns variability is a possible advantage of data-driven classification approaches. However, it needs to be demonstrated and applied with a minimum set of sensors for feasibility. To overcome feature selection limitations, [9, 6, 10] proposed different approaches based on Deep Learning, shifting the process from feature engineering to feature learning. They exploited the automatic feature extraction stage performed in different layers by Deep Neural Networks (especially CNN) on raw signal inputs to increase the robustness and performance of the models. Côté-Allard et al. [23, 17] and Rehman et al. [16], have shown that CNet architectures are able to correctly classify between six to eight hand movements. Furthremore, CNet architectures with transfer learning have shown promising capabilities to overcome EMG variability [23]. In the literature, there is still no agreement about the best set of features, or the use of deep learning on raw datasets, the optimal model, or neither about the electrodes' configuration setting: all these works are proposing different solutions based on heuristic assumptions and attempts. Furthermore, these models are difficult to interpret: the black-box nature of AI classification models is a significant obstacle for reflecting the achieved accuracy of classification back to the minimization of sensors. To this step, an effort to make the black-box approach explainable becomes mandatory.

2.2. Explainable AI

XAI can address these challenges and, at the same time, give physiological explanations in terms of how AI models have decoded EMG signals, producing details or reasons to make their functioning clear and easy to understand [12]. Post-hoc explanation techniques are based on a reverse engineering approach, i.e. opening the black-box model that has been already optimized for classification through independent XAI algorithms [24, 25]. These methods have the advantage of exploiting the high classification capabilities of not-interpretable

models while investigating their inner structure and the reasoning behind the black-box predictions. Post-hoc explanation techniques include those who provide feature importance [26, 27] and saliency masks [28, 29, 30]. XAI algorithms have been widely applied to images, text, and tabular data, but their use in real-world applications is still in a growing phase. In particular, XAI applied to physiological time series has been rarely investigated, although the topic strongly demands comprehensibility and interpretability. Some research groups have proposed to explain Electroencephalogram (EEG) time series, a brain signal that is recorded similarly to EMG, and they share common characteristics and challenges. Sturm et al. [31] is considered the benchmark for EEG interpretation and visualization. They proposed Layer-wise Relevance Propagation to explain a DNN trained for motor-imaginary Brain Computer Interface (BCI), highlighting the relevance of each input feature of EEG data using heatmaps. They obtained neurophysiological explanations, inferring the influence of visual activity and eye movements in the classification of motor tasks. Schirrmeister et al. [32]proposed two different approaches to interpret EEG CNN classification for motor tasks. They demonstrated that the network was reliable and learned EEG characteristics in agreement with domain knowledge. EEGNet [33] contributed with a comprehensive study about the best network architecture for EEG classification and proposed three different visualization approaches showing that neurophysiologically interpretable features can be extracted from the EEGNet model. Similarly, [34] assessed the importance of kernel dimension and filter values to produce band-pass filtering, showing that kernel learned by the model gives insights into the morphology and patterns of the input. Finally, [35] investigated amplitude changes, frequency bands, and phase changes at different layers of a network; the visualization helped to better understand the structure and characteristics of EEG time series. Recently, Côté-Allard and his group [17], have applied local saliency maps to Adaptive Domain Adversarial Neural Network for EMG classification to explain single EMG examples. Yu et al. [36], proposed a simple analysis based on principal components extracted directly from the activations of the last ReLU layer. These works support the applica-

tion of XAI as a tool for explaining neurophysiological phenomena, and their valuable contributions make XAI implementation for EMG pattern recognition promising.

3. The Proposed Approach

This Section provides an in-depth description of the proposed methodology for EMG classification and interpretation, summarized in Fig. 1. Step 1 is related to data collection and preprocessing. In Step 2, we perform hand gesture classification in two different ways. The first one is by means of "classical" machine learning methods, using 15 features obtained through a preliminary feature engineering phase. The second one is through deep learning, thus without a manual, expert-based feature selection. Those two methods are carried on in parallel because they provide different insights. On the one side, the machine learning models provide useful information on the importance of statistical and domain-related important features, in accordance to the literature, like, for instance, the average of the rectified signal or the number of times the signal crosses the zero. On the other side, the deep neural networks use directly the raw EMG signal, they provide important information on the topology of the problem and allow to infer highly plausible physiologically explanations of the classification results. Moreover, while deep learning methods show slightly better classification performance and promising solutions for addressing inter-session variability, machine learning methods provide, after the feature selection, good result in lower computational times, taking a step towards novel real-time applications. Even in Step 3, the XAI analysis, we interpret the classification results separately for machine learning and deep learning classification, aiming to obtain a deeper understanding of the importance of the hand-crafted classification features from the machine learning models as well as the importance of the hardware configuration from the deep learning models. Finally, in the final discussion, we bring together the two, proposing a global simplification of the EMG hand gesture classification problem both from a physical and a compu-

tational viewpoint. In addition, in the feedback steps, namely *Data collection analysis and error decoding* and *Transfer Learning for intersession variability* we present two additional benefits derived from the our XAI analysis.

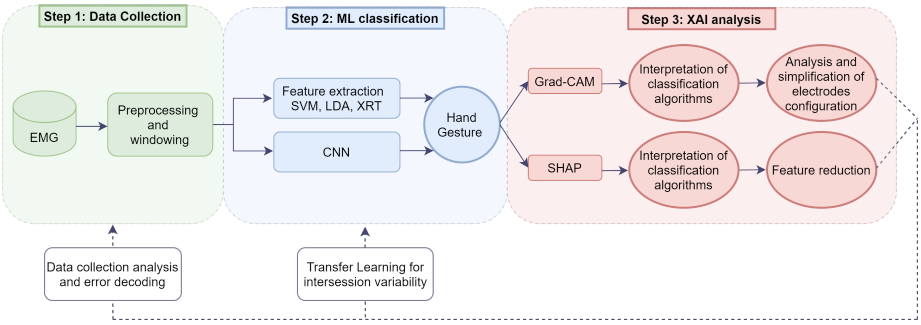


Figure 1: Pipeline of the proposed methodology for EMG classification and explanation

3.1. Experimental protocol and settings

The EMG data are collected within a real study [37] targeting different upper limb movements, following inclusion criteria according to the most frequent daily gestures. Each gesture has been performed starting from different hand starting posture, introducing variations to achieve higher robustness in pattern recognition. Labels for classification are defined as the hand gesture, independently from the starting postures. During the movement, the EMG activity of upper limb muscles has been recorded through 10 bipolar channels (surface electrodes) with a sampling frequency of $2048 Hz$. The electrodes were placed in two circumferences, one placed 3 cm from the elbow with six electrodes equally spaced around the circumference of the forearm, the other placed 3 cm distal to the previous one, , according to SENIAM (Surface EMG for non invasive assessment of muscles) guidelines [38]. Reference electrode was placed on the back of the wrist as suggested by SENIAM directions [38]. The position of electrodes is shown in Fig. 2. Despite the arrangement was geometrically defined, there is a rough a priori correspondence between the electrodes and the underneath muscles:

- Channel 1: flexor carpi ulnaris/flexor carpi radialis

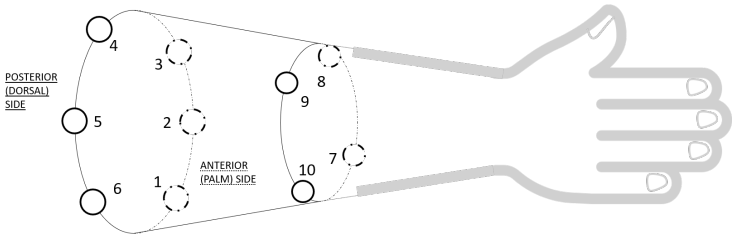


Figure 2: Schema of electrodes placement in the right forearm: six electrodes are placed equally spaced in a circumference close to the elbow. The remaining four electrodes are placed distally. The posterior part of the arm (dorsal) is in front of the viewer while the anterior (palm) side is on the back.

- Channel 2: flexor carpi radialis/ pronator teres
- Channel 3: brachioradialis / pronator teres
- Channel 4: ulnaris extensor carpi radialis longus
- Channel 5: extensor digitorum/extensor carpi ulnaris
- Channel 6: flexor carpi ulnaris
- Channel 7: palmaris lonugs/flexor carpi ulnaris/ flexor digitorum
- Channel 8: brachioradialis
- Channel 9: extensor carpi radialis brevis/extensor digitorum
- Channel 10: flexor carpi ulnaris

Then, raw EMG signals were collected and filtered using a $10 - 500$ Hz band pass filter (Butterworth order 4^{th}) and a 50 Hz Notch filter (Butterworth order 2^{nd})[39]. Furthermore, since the steady-state signal of muscles was targeted, 100ms from the beginning and end of windows were removed to eliminate the transient part of the movement signal. Aiming at online classification, segmentation is a relevant step for the delay in the response. The windows length was selected considering limitations for online applications. Finally, as suggested by Côté-Allard et al. [23], sliding window approach is the most effective augmentation technique for surface EMG classification. On the one hand, it increases the database size to achieve high performance and better generalization in deep learning methods. On the other hand, it mimics a real-time application where a

sliding window approach allows to reduce the delay between intended movement and classification, continously classifying the acquired EMG.

3.2. Personalized Models

EMG variability due to electrode positioning, anatomical differences, electrode displacement, and other inter-subject differences strongly limits the use of generalized models. Therefore, EMG classification often relies on personalized models. For each subject, his own data are used both to train and test. We classified hand movements with two approaches: classical machine learning, based on domain knowledge for feature extraction, and deep learning with convolutional neural networks. Each movement class has been divided into training (2/3 repetitions) and test (1/3 repetitions), equally balanced for hand starting postures. Accuracy, F1 score and roc-auc score were used as evaluation metrics.

3.2.1. Classical Machine Learning

When dealing with classical machine learning and time series, a typical approach is to window the signal and extract relevant features from each window, namely the feature engineering step. In an optimal scenario, the features are carefully selected to represent in a different dimensional space the same information that was present in the original signal. 15 different statistical and domain-based features, based on literature suggestions [40, 41, 4, 14, 23, 42], were extracted from each window, for each channel (Table 1). The formula of each feature is presented in Appendix A. These features were then grouped in two feature sets, Improved Time Domain (ITD) (MAV, ZC, SSC, WL, RMS, IEMG, HP_A, HP_M, HP_C) and Full Dataset (FULL) (MAV, ZC, SSC, WL, HP_A, HP_M, HP_C,SE,CC_1-4, RMS, IEMG, SKEW9). Then, we used this features set to train different machine learning algorithms: MLP, KNN, SVM, LDA and Random Forest, AdaBoost, Bagging Ensemble Trees and XRT. For each model, a grid-search approach has been applied to tune the hyper-parameters using the training set of a randomly chosen subject with 3 fold cross-validation. Each fold was created dividing different repetitions of each movement of the training set.

Table 1: Features extracted from EMG signal for classical machine learning classification.

Feature name	**Explanation**
Mean Average Value (MAV)	average of the rectified signal
Zero Crossing (ZC)	number of times the signals crosses the zero line
Slope Sign Change (SSC)	number of times that the signal slope changes
Waveform Length (WL)	a simple characterization of the signal's waveform, defined as cumulative length of EMG waveform over the time segment
Hjorth Parameter (HP)	statistical properties in signal processing HP_A: Activity HP_M: Mobility HP_C: Complexity
Sample Entropy (SE)	measure of the complexity of physiological time series
Cepstral Coefficients (CC)	coefficient of inverse Fourier transform of the log power spectrum magnitude of the signal (up to 4^{th} order)
Root Mean Square (RMS)	quadratic mean
Integrated EMG (IEMG)	sum of fully rectified signal
Skewness (SKEW)	measure of asymmetry of a distribution

3.2.2. Deep Learning

Differently from traditional classifiers, deep learning algorithms have the great advantage of learning effective data representations directly from the data, without the need for hand-crafted features [43]. Recent works using simple CNet architectures have shown promising performance for surface EMG-based gesture recognition using raw EMG directly as the input of the network [23, 16, 17]. In this work, the EMG input has been encoded as a 2D matrix with height equal

to the number of channels and width equal to number of samples in a window. It can be considered as a 2D grey-scale image where the axis are in the time-space dimension. The first session of DL training included two different CNN architectures: CNet1D and CNet2D.

Both architectures are designed with two main stages: the first stage is a sequence of convolutional blocks working as "feature encoding", and the second one is composed of fully connected layers for the final classification step. Each convolutional block consists of a convolutional layer with Rectified ReLU activation function, batch normalization, max pooling, and dropout. CNet1D has filters with shape $(1xW)$: it does not exploit the spatial relations between channels in the first dimension of the input (height) in the feature encoding stage, focusing on extracting information of individual channels only in time (width). On the other hand, CNet2D also performs space convolution applying a spatiotemporal convolution. It has the same architecture as CNet1D, but the filter size of each convolutional layer is $(2xW)$. In this way, the discrete information provided by single electrodes, which are juxtaposed in the first dimension (height) of the input, is merged and filtered together. In order to find the hyper-parameters of the networks, different combinations of hyper-parameters (number of layers, number of filters, filter size and activation function) were tested by trial and error on the validation set (1/5 of the training set). DL models have been trained with categorical cross-entropy loss function:

$$CE = -\sum_{c=1}^{C} y_{i,c} * \log \hat{y}_{i,c} \tag{1}$$

with $\hat{y}_{i,c}$ is the softmax probability for the c^{th} class and i_{th} instance. Then, the total loss is the sum of CE loss for each instance in the batch size. The total loss on the validation set was computed for early stopping. The best model on validation loss was saved as a check point. During the training phase of the proposed architecture, Adam Optimizer was used as an optimization method.

3.3. XAI approaches

In this article, we employ and adapt two different XAI methods. One is used to explain the classical machine learning classification fed with 15 handcrafted features typically used in this field, and to propose novel feature selection approach based on the explanations obtained. The second one is used to explain and visualize the activations that are relevant for the target movement in the neural networks' classification; this will allow us to propose a simplification of the hardware, maintaining statistically similar classification performances. For the first task we need, therefore, a global, feature-based explanation methodology that can be applied agnostically to several machine learning algorithms. As it emerges in one of the most recent surveys on XAI [13], among the global, model-agnostic XAI alogrithms, most of them provides rules as output, instead of feature relevance(see, e.g., [44, 45] and all the decision-tree based surrogate models), thus are not suitable for feature selection. Among the feature based ones, SP-Lime and K-Lime provide methods to select representative sets by means of submodular optimization [26], while SHAP provides global explanations considering all the set of instances on which the model has been trained based while ensuring several property derived from probabilistic game theory [27, 46, 13].

The second task requires a method for post-hoc visual explanations of CNN classifiers. Several relevant methods have been proposed in literature. Some of them, despite producing fine-grained representations,are not class-discriminative [47]. Others, based on attention, needs to retrain the model adding attention mechanisms (see e.g. [30, 48]), which is our case is not doable, because, as common in post-hoc explanations, we explain an already trained black-box model [37]. Among the high-resolution, class-discriminative, post-hoc visual explainers, Grad-CAM has shown to achieve the best heatmap localization ability [29, 49] and have shown promising result in EEG-classification [50, 51]. Furthermore, another advantage of Grad-CAM is that it can be implemented on any CNet-based architecture without requiring re-training.

3.3.1. SHAP on classical machine learning

SHAP aims to explain the prediction of an instance x by computing the contribution of each feature to the prediction, exploiting Shapley values from coalitional game theory. SHAP belong to the category of methods referred to as *Additive feature attribution methods*. Additive feature attribution methods provide an explanation model that is a linear function of binary variables:

$$g(z^{'}) = \Phi_0 + \sum_{i=1}^{M} \Phi_i z_i^{'} \tag{2}$$

where $g()$ is the explanation model, $z^{'}$ is an interpretable representation of a simplified input $(x_i^{'}) \in \{0,1\}^M$, M is the number of input features and Φ_i the effect of each feature $\in \Re$. The idea in Eq. 2 is that summing the effect of all features in a linear explainable model is possible to achieve an approximation of $f(x)$, the output of the original opaque model. SHAP [27] provides a solution to Eq. 2 that respects three desirable properties: local accuracy, missingness and consistency. SHAP values, i.e. each feature contribution, of the instance x, can be computed as follow:

1. Sample coalitions $z'_k \in \{0,1\}^M$ (1 feature present, 0 feature absent), with $k = 1...K$ different coalitions

2. Converting z'_k in the original space using $h_x(z)$; if the feature is present (1) its value is equal to the value in the original instance x, if the feature is absent its value is randomly sampled by the marginal distribution; find the prediction $f(h_x(z'_k))$ with the original opaque model

3. Compute the weight Φ_k for each z'_k

4. Fit weighted linear model $g(z')$ (Eq. 2) minimizing a loss functions weighted depending on the number of present features in the coalition

5. Return Shapley values Φ_k the coefficients from the linear model

The solution of the linear model is obtained minimizing the following loss function:

$$\mathcal{L}(f, g, \pi_{x'}) = \sum_{z' \in \mathcal{Z}} [f(h_x(z')) - g(z')]^2 \pi_{x'}(z') \tag{3}$$

$$\pi_{x'}(z') = \frac{(M-1)}{\binom{M}{z'}|z'|(M-|z'|)} \tag{4}$$

where $|z'|$ is the number of non-zero elements in z'. The distance kernel in Eq. 4 aims to return largest weights for small coalitions (i.e. few 1's) and large coalitions (i.e. many 1's). The idea is that from small/large coalitions it is possible to learn more about individual features since their effect is studied isolated from the others [52].

We computed SHAP values independently for each instance resulting in $SHAP_{scnf}$ (with s subjects, c classes of movements, n number of observations for each subject s and f features).

SHAP for Feature Selection.

We proposed a **correlation-XAI approach for feature selection** merging generalized feature importance obtained with SHAP and correlation analysis to achieve the optimal feature vector for classification. The selection of the best set of features is problem-specific, and it is constrained by feature significance, reliability, robustness, complexity and maximum class separability [53, 54]. A common agreement about the best set of feature has not been yet achieved and feature selection is usually carried out with sub-optimal unsupervised approaches. In our approach, we firstly computed correlation matrices of all features, dropping features with correlation higher than 0.9. Then we computed:

$$\mu_{SHAP,cf} = \frac{1}{S} \sum_{s=0}^{S-1} \frac{1}{N_s} \sum_{n=0}^{N_s-1} |SHAP_{scnf|} \tag{5}$$

and in the same way the variance σ_{SHAP}. The average SHAP value $\mu_{SHAP,cf}$ was an index of the absolute importance and contribution of the feature f for the prediction of class c, i.e. the gesture, while the variance represented its consistency. Following a deep analysis of scatterplots representing $\mu_{SHAP,cf}$

and σ_{SHAP}, we highlighted five different possible set of features and used each of them to re-train the models. These sets were:

- Original FULL set
- Original ITD set
- Not-correlated features obtained after the correlation analysis
- *Best* set, obtained after correlation-XAI analysis
- *Worst* set, features identified as not important with SHAP

We compared performances and validate the results with Wilcoxon Signed-Rank test. Statistical significance was set to 0.05.

SHAP for EMG Variability.

The variability of EMG features due to electrode shift, different arm position, changes in arm posture and fatigue among other causes, is an important characteristic for the robustness and reliability of a model. We exploited SHAP values to understand both qualitatively and quantitatively whether features were robust in **time, inter-posture, inter-session or inter-subject**. We visualized the results of EMG variability analysis in heatmaps representing feature importance according to Eq. 6:

$$SHAP_{nf} = \frac{|SHAP_{nf}|}{\max(SHAP_{nf})} \tag{6}$$

We computed the absolute value of Shapley values and then normalized it, resulting in $S \times C \times P$ matrices of SHAP values (S number of subjects, C number of classes-gestures, P number of postures). Firstly, we evaluated feature importance in time, comparing SHAP values for different windows of the same movement evaluating the robustness of the features. The second step was to compare heatmaps obtained for the same subject but in different hand starting postures. Thirdly, an inter-session analysis was conducted for two subjects whom performed the experiment twice. Finally, an inter-subject analysis was performed to evaluated differences in the electrodes placement and in

the anatomy. The qualitative observation of these heatmaps was supported by Kruskal-Wallis tests.

3.3.2. Grad-CAM for CNNs

Grad-CAM [29] is a model-specific, outcome explanation method based on saliency maps. It proposes a class-discriminative localization technique for CNN interpretation that outputs a visualization of the regions of the input (heatmaps) that are relevant for a specific prediction. Firstly, the gradient of y^c (score for the class c) is computed with respect to each feature map activation A^k, $\frac{\partial y^c}{\partial A^k}$, with k index of the specific feature map. Then $\frac{\partial y^c}{\partial A^k}$ are global average pooled to obtain the weights α_k^c:

$$\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k} \tag{7}$$

with Z number of pixel in the feature map and i and j are indexing along width u and height v of the feature map ($\frac{\partial y^c}{\partial A_{ij}^k}$ refers to the activation at location i, j of the feature map A^k). α_k^c represents the importance of the feature map k for the class c. Then the class-discriminative localization map Grad-CAM $L_{Grad-CAM}^c \in \Re^{uxv}$ is computed as:

$$L_{Grad-CAM}^c = ReLU(\sum_k \alpha_k^c A^k) \tag{8}$$

returning the importance of each neuron of the last convolutional layer. Lastly, $L_{Grad-CAM}^c$ is upsampled to the input image dimension using linear interpolation, returning the importance of each pixel of the input image. Selvaraju et al. [29] proposed Grad-CAM as a local XAI model that can be applied to image classification (e.g. with ImageNet weights), image captioning, and visual question answering (VQA). These examples require proper images that are different from EMG signals, i.e. dynamical physiological time series characterized by time-space dimensions. Furthermore, EMG signal is characterized by high dimensionality (e.g. sampling rate), channel correlation, artifacts, and noise that are not typical of static images, making learning features in CNN layers and extracting explanations with XAI, more difficult for EMG signals than for

classical images. On the other hand, EMG signals analyzed as a 2D image have the advantage to present similar pattern in different windows, since each channels is always placed in the same position. In this work, starting from the *local explanations* computed with the original algorithm [29], we propose a generalization to obtain *global explanations*. Firstly, we compute Grad-CAM saliency mask for each instance as in Eq. 7 and 8, obtaining $L_n^{s,c}$ feature importance maps with n instance $\in \{1...N_c\}$, c class $\in \{1...C\}$ and s subject $\in \{1...S\}$. $L_n^{s,c}$ has dimension $ChxT$ that is the width (time) and height (channel) dimensions of input image. Then, each saliency maps is normalized:

$$L_n^{s,c} = \frac{L_n^{s,c} - min_{ch,t}(L_n^{s,c})}{max_{ch,t}(L_n^{s,c}) - min_{ch,t}(L_n^{s,c})} \tag{9}$$

For each $L_n^{s,c}$ we compute the average value in time (from $Ch * T$ to a vector $Ch * 1$). Then, we averaged all vectors importance for each subject and for each class weighting depending on the value of the prediction $\hat{y^c}$:

$$L_{ch}^{s,c} = \frac{1}{N_c} \sum_{n=0}^{N_c-1} \hat{y}_n^{s,c} \left[\frac{1}{T} \sum_{t=0}^{T-1} * L_{n,(t,ch)}^{s,c} \right] \tag{10}$$

$L_{ch}^{s,c}$ represent the average importance of each channel ch for the prediction of the class c for subject s. In order to analyze the achieved results, we firstly propose an input feature importance visualization based on the local explanations (Eq. 8) aiming at analyzing relevant input features that have led to gesture classification and how they are connected with physiological characteristics. Local explanations are plotted as a 2D heatmap (time x space). Besides heatmaps, we plot the original EMG signals with Grad-CAM explanation superimposed to analyze the dynamics of the signals.

Grad-CAM for reducing hardware and software complexity.

The usability of upper limb prostheses is a significant obstacle to prostheses application and is strongly influenced by the number of electrodes [3, 22]. Following results and observations that will be presented in Section 4.4, we select the optimal number of electrodes. Four Convolutional Neural Network models have been designed to classify hand movements using only six channels.

The architecture choice has focused on 2D convolution to compensate for the information loss with 2D filtering in space. Then, we have further validated the new set of electrodes, harnessing them to address another challenge related to myoelectric prostheses, i.e. inter-session variability [8, 16]. The aim of myo-controlled prostheses is not the laboratory but real-world conditions, where day-to-day variations might strongly influence the classification performance. To this end, we proposed a continuous learning approach with preliminary experiments on the inter-session variability between two different independent measurement sessions of the same subject. Continuous learning is a popular approach, especially in AI-powered devices and IoT [55, 56], defined as the ability of a model to learn continually from a stream of data. This can be done by triggering a learning phase periodically. Myo-controlled prostheses is a suitable field for continuous learning: it enables to collect more data increasing the classification performance progressively, and at the same time proposing a solution for day-to-day variability. Generally, continuous learning can be done both retraining the model from scratch or tuning a pre-trained model on new data, minimizing the time required for the training phase. We tested our hypothesis on a subject that has performed hand movements in two independent measurement sessions. The subject's first session was used to train a CNN model with six channels; the best model on the validation set was selected, and the weights of the first two convolutional blocks were frozen. Then, the data from the second session were used to fine tune latest convolutional and fully connected layers.

4. Experimental Results

In this Section, we briefly introduce the results of the best performing models (SVM, LDA, XRT, and CNN) on EMG signal decoding. A extensive analysis of the other machine learning methods used in this work is presented in Appendix B. Then, we proceed with SHAP and Grad-CAM with an in-depth domain knowledge analysis of the explanations achieved towards real-world applications.

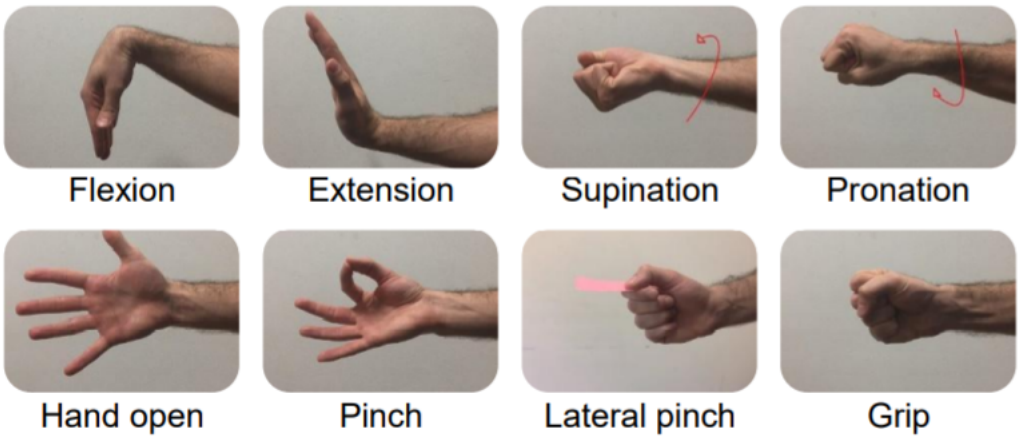


Figure 3: Representation of the 8 targeted hand movements

4.1. Experimental Study Design and Settings

The experimental study involved 11 subjects, 5 females and 6 males, with no history of neuromuscular diseases; two subjects repeated the experiment twice. Among hand movements, eight gestures have been targeted: wrist flexion, wrist extension, supination, pronation, hand open, pinch, lateral pinch, grip/fist. The protocol included all hand gestures performed from three different starting postures: palm faced upward, palm faced sideways, palm faced downward. In each hand starting posture, each gesture was repeated 5 times, resulting in 8 gestures $\times$5 repetitions $\times$3 starting postures = 120 total movements. For each movement, the 15 repetitions were divided into training (2/3) and test (1/3) set; then, the training repetitions were further divided into training and validation set. Each gesture was performed for 5 seconds. Considering that for online applications window length plus processing time to generate classified control commands should be less than 300ms [57], window size of 250 ms was selected, with a stride of 62.4ms, resulting in overlap among consecutive windows of 187.5ms.

4.2. Personalized Models Classification Results

4.2.1. Classical Machine Learning

We selected 8 ML apporaches: KNN, MLP, SVM, LDA, XRT, Bagging, Adaboost e Random Forest. For each of them we performed a grid-search on several hyperparamenters, to maximize Accuracy and F1 score, similarly to what have been done in [58] and [59]. The models have been validated using 3-fold cross-validation. For sake of reproducibility, the random seed has been fixed. The full set of hyperparameters and the classification results are reported in Appendix B. LDA with svd solver, SVM with regularization=1, a linear kernel, and $\gamma = 0.01$ and XRT with 50 estimators, have produced the best performance, shown in Table 2. Best LDA (with FULL feature set) and best

Table 2: Classical Machine Learning results in term of average accuracy (%), f1 score (%) and roc-auc sscore (%) with inter-subject standard deviation

	Accuracy		F1-score		Roc Auc score	
Classifier	Improved TD	FULL	Improved TD	FULL	Improved TD	FULL
LDA	91.62 ±4.18	92.55 ± 4.23	91.69 ±3.95	92.60 ± 4.02	95.22 ± 2.28	95.75 ± 2.31
SVM	91.72 ± 3.60	91.47 ±3.42	91.70 ± 3.38	91.44 ±3.21	95.23 ± 1.93	95.11 ± 1.86
XRT	90.12 ± 3.85	90.02 ±3.70	90.13 ± 3.73	89.93 ±3.71	94.40 ± 2.13	94.38 ± 2.15

SVM (with ITD feature set) showed no statistical difference ($pval = 0.248$). XRT was statistically different from LDA ($pval = 0.013$) but not from SVM ($pval = 0.074$). However, we considered XRT models because SHAP [60] is considerably faster in computing feature importance on them with respect to other models.

4.2.2. Deep Learning

The final CNet1D and CNet2D's architectures are shown in Appendix C. Both model are designed as a CNet architecture with three convolution blocks followed by two fully connected blocks with a final softmax layer for classification. Each convolutional block includes convolution, batch normalization, RRelu activation function, max pooling and drop-out. The main difference is that CNet1D applies filtering only in the time dimension (along the width of

the input) and it does not exploit the relations between channels while CNet2D performs a 2D filtering both in time (width) and space (height). The filters parameter (number of filters, kernel size and stride size), as well as the max pooling and dense parameters have been optimized with a trial-and-error strategy. Final architecture hyperparameters are shown in Appendix C. Adam optimizer has been used as optimization strategy. The final learning parameters for CNet1D and CNet2D's training are learning rate from 0.003 to 0.0005 with a factor of 0.5 and patience 30, 400 epochs with early stopping patience 100 (based on validation loss) and batch size equal to 128. The preprocessed raw data was passed directly as an image of shape 10512 (Channel $\times$ Samples) to the CNet. The final input size was (batch size x 10 x 512 x 1). CNet1D and CNet2D have been trained for each subject achieving an overall average accuracy with variation among subjects of **93.23% $\pm$ 2.77** and **92.81% $\pm$ 3.06** respectively. We have tested both classical machine learning and deep learning models, with the latter showing better performance. Therefore, we decided to focus on CNNs, applying XAI for models and outcomes explanation to better understand their hidden functioning.

4.3. SHAP

SHAP values were computed independently for each instance of all 11 subjects (and models), resulting in $SHAP_{scnf}$ (with s subjects $\in \{1...11\}$, c classes of gestures $\in \{1...8\}$, n number of observations for each subject $s \in \{1...N_s\}$ and f features $\in \{1...F\}$). Each SHAP value represented the importance of the feature f for the prediction of class c for the instance n of the subject s.

4.3.1. Correlation-XAI approach for feature selection

In Eq. 5, we have introduced a generalization of SHAP local explanations merged with correlation analysis for feature selection. We plotted μ_{SHAP} and σ_{SHAP} (Eq. 5) in scatterplots (an example is shown in Figure 4).

Correlation-SHAP analysis through scatterplots highlighted RMS, HPM and HPA as the most important features. In particular, RMS was relevant for all

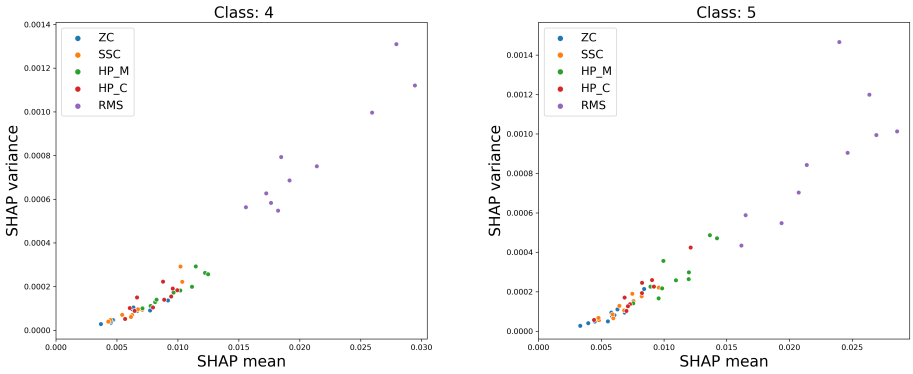


Figure 4: Scatterplot SHAP values for SVM class 4 (hand open) and 5 (pinch) obtained after applying SHAP to SVM models trained without correlated features.The x axis represents μ_{SHAP}, y axis σ_{SHAP} and the hue (color of different points, that can display a third variable) summarizes the feature typology

classifiers and all classes, as evident by its positioning on the right side of the scatterplots (Figure 4) This result is coherent with domain knowledge: indeed, a stronger electrical activity is linked to a higher EMG signal and thus, to the contraction of specific muscle fibers during the movement.

Thereafter, we trained ML models with five different feature sets to validate our approach; the performances of SVM are shown in Figure 5 in term of accuracy and F1-score, and Wilcoxon Signed-Rank test is presented in Table 3. Similar resulst were achieved for the other classifiers. A broad analysis of SHAP

Table 3: Wilcoxon signed-rank tests for SVM with different feature sets: FULL, ITD, features after correlation analysis, best set with best features obtained through correlation-XAI feature selection and worst set with features discarded by SHAP

		ITD	**Not Corr**	**Best**	**Worst**
FULL	pval	0.722	0.013	0.508	0.003
ITD	pval	-	0.091	0.722	0.003

results highlights that:

- Correlation feature selection alone may not lead to the best set of features: Not correlated features show worse results than FULL features set ($pval <$

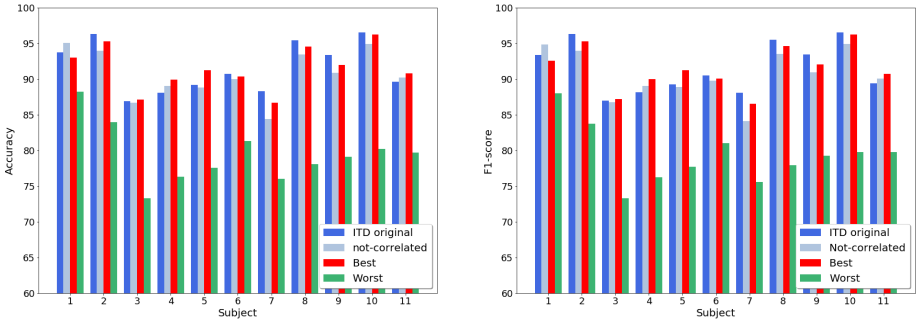


Figure 5: SVM performances in correlation-XAI feature selection, in term of accuracy and F1-score

0.05)

- The best set of features, i.e. the one obtained combining correlation and SHAP (RMS, HPM, HPC), achieves competitive performances comparable with the models trained both with FULL and ITD features ($pval >> 0.05$) while decreasing the inference time from 275 ms to 0.55 ms.

- The features identified by XAI as not important (ZC and SSC) are indeed significantly worse that all the other feature sets.

- Features are robust among classifiers: SHAP achieved similar feature relevance for SVM, LDA and XRT. The importance of these features do not depend on the specific classifier, supporting their significance in the specific domain.

Limiting the number of features has benefits not only for inference and training time but is a great advantage to overcome the curse of dimensionality [61]: with a large number of features, i.e. large hypothesis space, the variance of the model increases requiring a larger dataset for a correct estimation of the prediction error.

4.3.2. EMG Variability

The evaluation of time, inter-postures, inter-session and inter-subject variability in EMG pattern recognition for robust myo-electric control, has been carried out through a qualitative- quantitative analysis of SHAP features importance. We firstly group $SHAP_{scnf}$ by hand starting posture, and then we plot features and instances in heatmaps; an example is shown in Figure 6.

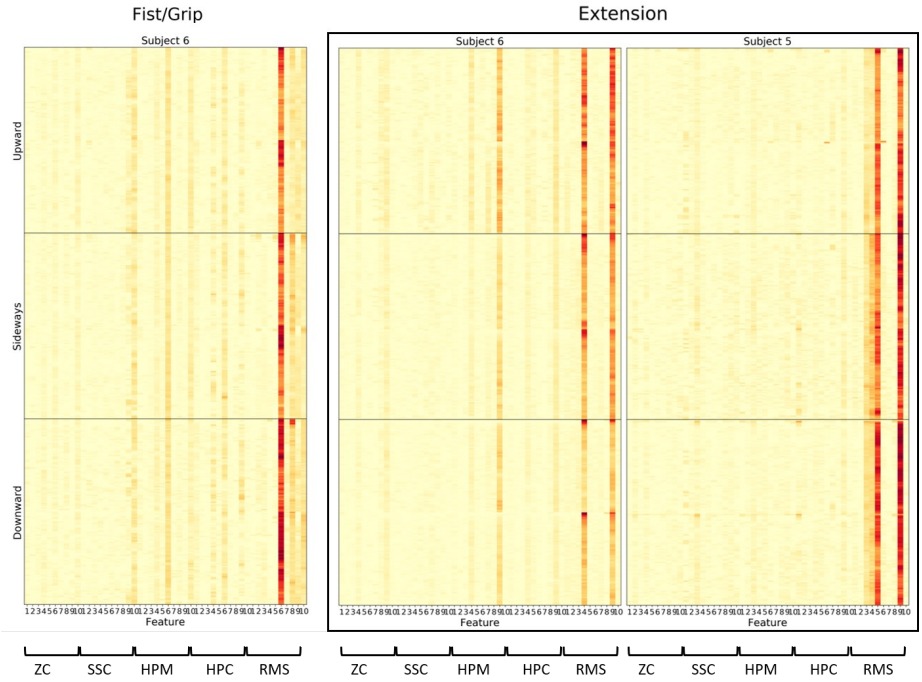


Figure 6: SHAP values for EMG variability. For each heatmap: *x axis*: features obtained from the correlation analysis in the following order: ZC, SSC, HPM, HPC, RMS; each one for 10 channels; *y axis* represents the time step (different instances are ordered in time); color intensity changes according to the importance of the feature for the prediction (Eq. 6). Each row is a different starting posture of the hand (upward, sideways and downward). The first column shows fist/grip gesture for subject 6; the last two columns present the extension gesture for subject 5 and 6.

We can notice that RMS, the last ten columns, has a significantly higher intensity than the others, supporting the results achieved in Section 4.3.1. We observe that features importance appears to be consistent in time and posture. It

is possible to evaluate the time dimension looking at the trend of each column in the y axis. These heatmaps do exhibit a predominantly vertical pattern, where the intensity of each point remains essentially uniform. Similarly, contiguous figures present similar patterns; for instance, in Figure 6, RMS of channel 6 is the most important feature and its value is almost constant in time and among rounds. Qualitative observations have been validated through Kruskal-Wallis statistical tests. **Time and inter-posture variability is not significant** since the comparison of different heatmaps shows similarity of feature importance in time and in respect to different hand starting postures, supported by Kruskal-Wallis $pval >> 0.5$. Therefore, electrode shift, short-time skin conductivity changes, hand starting posture and the orientation with respect to gravity are not relevant factors that are affecting the classification performances. The system is robust in short-time and against different starting position as usually assumed in real life clinical applications.

On the other hand, inter-session (multiple independent acquisitions) and inter-subject variability is significant, validated through Kruskal-Wallis statistical tests that showed significant statistical differences. We observe that generally, what is changing between subjects or sessions is not the feature type (e.g. RMS) but the channel from which it is extracted, especially among close/ adjacent electrodes. For instance, Figure 6 shows that for subject 5 RMS of channel 4 and RMS of channel 9 are more important while for subject 6 RMS of channel 5 and RMS of channel 9 dominates. In Figure 2, we could notice that electrode 4 and 5 are adjacent; this is suggesting that the inter-session or inter-subject variability is mainly due to slightly differences in the electrode placement or anatomy of the subject. Improving the electrodes' placement with more robust protocols should improve inter-session or inter-subject performances.

4.4. Grad-CAM

We applied Grad-CAM on the test set classified using CNNs, following the steps introduced in Section 3.3.2. Tab. 4 shows a global explanation in terms of the most relevant channels for each hand gesture prediction. We linked general-

Table 4: Most important electrodes for each gesture, obtained ordering channels in respect to the averaged Grad-CAM values for all instances and subjects. Only the best three electrodes for each gesture are presented. In the last row, $L_{ch}^{s,c}$ in Eq. 10 have been further averaged among subjects and the correspondent most importance channels are shown.

	Gesture							
Subject	Flex	Ext	Sup	Pron	Open	Pinch	Lat Pinch	Fist
1	1,7,10	1,9,7	7,1,3	1,2,6	10,3,4	8,4,1	7,1,2	8,7,6
2	1,10,6	8,4,9	7,3,5	4,8,3	4,3,1	7,4,10	7,1,10	8,7,6
3	6,3,1	4,9,3	9,10,3	4,10,9	10,6,3	10,6,3	10,2,9	8,4,3
4	1,10,3	9,4,5	3,2,5	4,8,3	1,3,2	10,4,8	3,1,2	9,10,6
5	4,1,8	9,4,10	4,8,9	4,8,2	10,7,6	10,6,3	2,8,4	6,10,7
6	8,1,10	9,5,4	5,10,2	10,9,8	8,3,7	2,8,7	10,5,9	10,8,6
7	7,1,10	9,8,4	5,7,3	6,5,3	9,4,7	3,8,4	3,8,7	8,6,3
8	5,9,8	3,9,4	3,2,8	8,10,3	1,2,10	8,3,2	9,6,5	6,7,1
9	7,10,1	9,4,5	4,8,9	2,3,8	10,3,7	8,3,6	1,8,2	7,1,4
10	5,1,10	4,9,5	8,6,5	3,7,8	10,6,3	4,9,8	8,7,5	7,1,10
11	3,10,1	9,4,5	7,3,9	8,3,4	5,9,4	5,4,3	10,7,3	7,1,6
ALL	1,10,7	9,4,5	3,4,8	4,10,3	10,3,4	8,3,4	7,10,2	6,7,8

ized Grad-CAM results with domain knowledge about myo-controlled prostheses and biomechanics of hand movements gaining the following insights.

4.4.1. Simple/complex movements provide less/more variability in channels' importance.

We can assume that the simplicity/complexity of a hand movement depends on the number and position of active muscles and actively involved joints, reflected in the noise and interference present in the recorded sEMG signals. Indeed, some movements such as wrist flexion and wrist extension are a single joint task carried out by specific, identified superficial muscles. For these movements, the superficial positions of the muscles, located close to the skin where the surface electrodes are placed, reduces the possible noise and artifacts. These movements can be defined as "simple" since, from a point of view of recording and pattern recognition, they should be always identified and classified with the same signals. Indeed, these movements (Tab. 4), especially wrist flexion and wrist extension, showed high robustness of channels' importance among

subjects: for example, for almost all subjects, hand extension is identified employing channels 9, 4, and 5. On the other hand, more complex movements such as supination, hand open and "finger movements" such as pinch and lateral pinch require either the use of deeper or thin muscles (e.g. supinator muscle is really profound); they provide a noisy signal in other channels due to movement artifacts or anthropometric differences. Thus, the variability of channels' importance is higher since each model may have learned how to identify these complex movements in a slightly different way.

4.4.2. Channels' importance identifies areas rather than single electrodes.

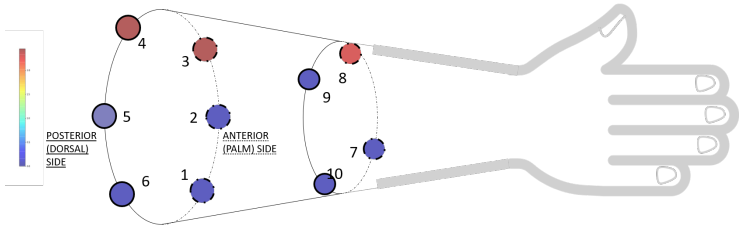


Figure 7: Schema of important electrodes for supination and pinch. Electrodes 3,4,8 are placed close to each other, 3 and 4 in the proximal circumference, 8 in the distal one

In Tab. 4, we can notice that channels' importance identifies areas of the forearm rather than single electrodes. For instance, supination and pinch are classified exploiting mainly electrode 3,4,8. In the physical electrodes' setting, electrodes 3 and 4 are placed adjacently in the proximal circumference while electrode 8 is placed distally between them (Fig. 7). Similarly, wrist flexion and wrist extension are classified with electrodes 1,7,10 and 4,5,9, respectively; these channels are placed in contiguous areas in the forearm.

A movement is usually carried out by the contraction of several functional synergic muscles, and each of them is active in its entire surface and not only in the specific point where an electrode samples it. Our results support that CNNs have automatically learned the correlation among close channels placed on the same muscle, fusing the information of functional synergic muscles in agreement with domain knowledge. On the other hand, this raises questions

about the redundancy of all channels, possibly leading to system simplification. For instance, channel 5 and channel 9 are both placed on a superficial area that roughly corresponds to the extensor digitorum muscle; it is reasonable to assume that by removing one channel, the model can still extract the relevant information from the remaining one. To validate these assumptions, we will test (Section 4.5) whether the CNNs are still able to learn relevant information and maintain high performances even dropping some redundant channels.

4.4.3. A deep analysis of wrist flexion and wrist extension shows that the model is learning meaningful and relevant characteristics in agreement with domain knowledge.

Following the division of the movements into simple and complex, we have decided to investigate wrist flexion and extension extensively.

Wrist flexion is carried out by a group of muscles called flexors, including superficial muscles (flexor carpi ulnaris, flexor carpi radialis, flexor digitorum) and other deeper muscles. In Tab. 4, we have highlighted that our model identifies wrist flexion through channels 1, which is placed superficially in correspondence of the flexor carpi ulnaris and the flexor carpi radialis, 10 the distal part of flexor carpi ulnaris, and 7 the flexor digitorum and the palmaris longus.

Therefore, our models have learned physiological relevant characteristics of the input to classify wrist flexion, automatically highlighting the flexors' electrodes as the most important channels to extract information for this specific class. Fig. 8e shows a saliency mask obtained applying Grad-CAM to one window labeled as wrist flexion. Fig. 8a represents the normalized average value for each electrode in a schematic representation localized on a forearm schema, to highlight the activity of the flexors.

We can notice that electrodes 1 and 10 are characterized by similar feature importance simultaneously, 2,6,7 are partially contributing, and 3,4,8,9 are not employed at all. Furthermore, the feature importance is not spread in time, but it is focused only on a specific time range of the signal. To further investigate the peculiarity of wrist flexion, time series signals are plotted and their impor-

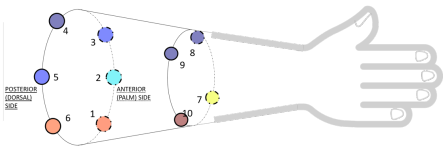
tance has been superimposed with colored markers in Fig. 8e. Firstly, EMG signals present larger amplitude in electrodes placed on flexors' area and larger Grad-CAM values correspond to high amplitude EMG activity, coherently with domain knowledge. Furthermore, in Section 4.3.1, we have shown that RMS is the most significant feature for machine learning classification. RMS is defined as the square root of the average power of the EMG signal for a given period of time. Thus, our results show agreement among manually extracted features and CNNs.

Wrist extension is carried out by a group of muscles, both superficial and deep, called extensors. Channels 4, placed on the extensor carpi radialis longus, 9, on the extensor carpi radialis brevis and extensor digitorum, and 5, on the extensor digitorum and the extensor carpi ulnaris, are mainly contributing to the classification of wrist extension (Tab. 4). CNNs are learning relevant features and automatically detect electrodes placed on extensors' muscles as the ones that contribute the most to wrist extension classification. Fig. 8d presents a heatmap for wrist extension; Fig. 8b represents the average value on a forearm schema; Fig. 8f further supports that signals with high amplitude and large oscillations are relevant for the prediction. These signals correspond to electrodes placed on extensor muscles, the physiologically activated muscles for wrist extension.

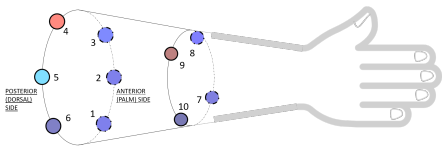
In this context, our explanations have a great value, verifying that CNNs automatically learn channels and properties of EMG signals relevant to the prediction as expected from the state of the art in the field. Thus, the classification is not performed based on some random noise or artifactual components but on meaningful characteristics of hand gestures. Furthermore, the correspondence found between oscillations and the importance of different channels further supports the redundancy of the information provided by all electrodes.

4.5. Insights towards optimized experimental solutions

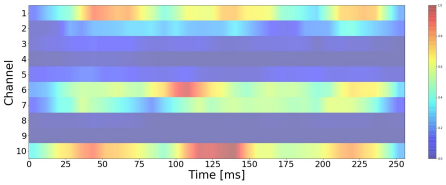
This section presents the optimization of EMG electrode configuration following the insights derived through XAI analysis. We trained four CNNs (Tab. 5) using as inputs only the six proximal EMG channels while the four distal



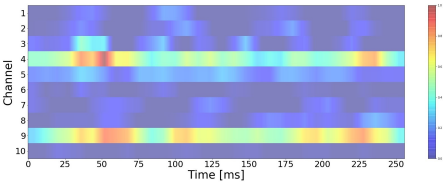
(a) Wrist Flexion - Forearm schema



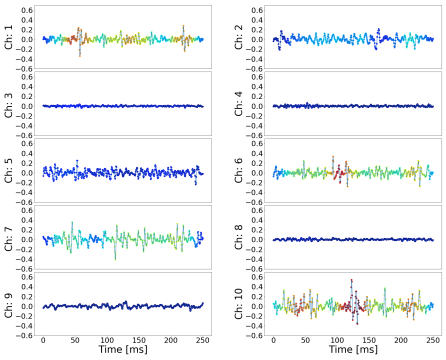
(b) Wrist Extension - Forearm schema



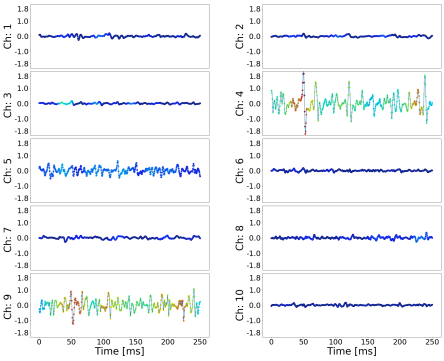
(c) Wrist Flexion - Saliency heatmap



(d) Wrist Extension - Saliency heatmap



(e) Wrist Flexion - EMG signals



(f) Wrist Extension - EMG signals

Figure 8: Grad-CAM results - Wrist Flexion (left) and Extension (right). In Fig. a the electrodes placed on flexors muscles are mostly activated, while Fig. 8b presents a more intense activation for muscles on the extensor, correctly identifying muscles' synergies. Saliency heatmaps (Fig. 8c and 8d): *x-axis*, time (total length=512 samples), *y-axis,* channels. The color intensity of each pixel is proportional to the importance of the specific channel-time point. Channels 1,6,7,10, placed on flexor carpi radialis, flexor carpi urnali e flexor digitorum are the most important channels for flexion, while channels 4 and 9, placed on extensor carpi radialis longus, and extensor carpi radialis brevis and extensor digitorum are the most important for extension. Finally, in Fig. 8e and 8f, feature importance is proportional to larger amplitude activity and oscillations.

have been removed. Our choice to keep the proximal sensors instead of randomly

Table 5: CNet architectures with 6 channels input; for all convolutional layer padding=same, while for maxpooling layers padding=valid.

	Layers	**CNet2D_3L**	**CNet2D_2L**	**CNet_2D_2L_big**	**CNet_1D_2D**
1	Input				
	Conv	(2x13)x32	(2x15)x32	(2x15)x32	(1x15)x32
	BatchNorm				
	RReLU				
	Maxpool	(1x4), s=(1x4)	(1x4), s=(1x4)	(1x4), s=(1x4)	(1x4), s=(1x4)
	Dropout	0.5	0.5	0.5	0.5
2	Conv	(2x9)x48	(2x15)x48	(2x30)x48	(2x30)x48
	BatchNorm				
	RReLU				
	Maxpool	(1x4), s=(1x4)	(1x4), s=(1x4)	(1x8), s=(1x8)	(1x8), s=(1x8)
	Dropout	0.5	0.5	0.5	0.5
3	Conv	(2x5)x64	-	-	-
	BatchNorm		-	-	-
	RReLU		-	-	-
	Maxpool	(1x4), s=(1x4)	-	-	-
	Dropout	0.5	-	-	-
4	Flatten			-	
	Dense	50, act=ReLU	50, act=ReLU	50, act=ReLU	50, act=ReLU
	BatchNorm				
	Dropout	0.5	0.5	0.5	0.5
	Softmax	8	8	8	8

sampling six out of ten was driven by the fact that amputees generally have a smaller - still intact and functioning - skin surface compared to healthy subjects, thus selecting the most proximal electrodes increases the amputees' population suited for this system for real-world applications [3, 6]. In Fig. 9, we compare CNet1D trained with 10 channels (average accuracy 93.23%), CNet2D trained with 10 channels (average accuracy 92.81%) and CNet1D_2D trained with only 6 channels (average accuracy 93.07%); there are not significant differences between the three models (Wilcoxon Signed-Rank $pval >> 0.05$). Competitive results are also obtained with all the different architectures proposed in Tab. 5 (92.34%, 92.61%, 93.02% respectively), supported by Wilcoxon Signed-Rank statistical tests ($pval >> 0.05$). These results further validated the hypothesis

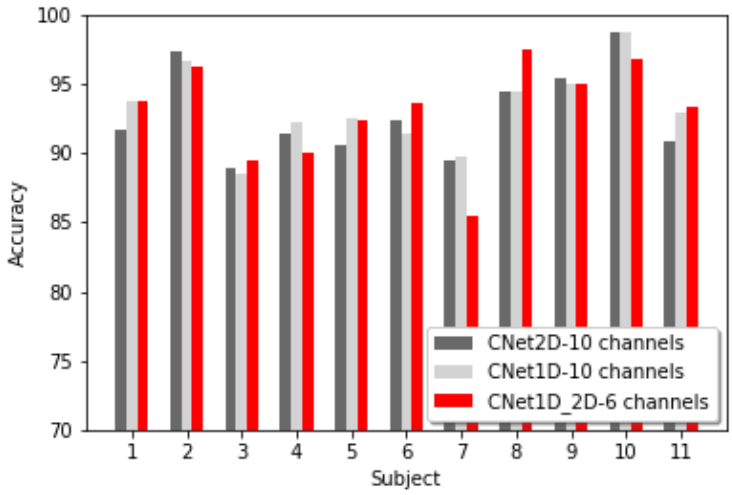


Figure 9: Comparisons of CNNs results, original models trained with 10 channels and one model, CNet1D_2D, trained with 6-channels input, in terms of accuracy. Wilcoxon signed-rank test supports not significant differences between the models (CNet1D-10 vs CNet1D_2D-6 pval=0.929 , CNet2D-10 vs CNet1D_2D-6 pval=0.594)

of redundancy among channels, identified with Grad-CAM explanations in 4.4. We have optimized the EMG electrodes' configuration reducing the number of superficial electrodes to six while maintaining the same performances of using a complex setting with ten channels. Besides achieving competitive classification performances, the benefits of the new setting with six proximal electrodes are several. Firstly it requires only six electrodes, reducing the complexity and cost of the hardware, thanks to a lower number of electrodes and simpler conditioning circuits also minimizing the risk of failure of the system. Then, it reduces the computational time and resources needed for training and inference, ensuring lower delay in real-time applications. The inference time of the original network was $15ms$ while the inference time of new networks is $\sim 10ms$, without considering the time for sending and processing the data that decreases for smaller input data. In particular, we minimized the required body surface for electrode placement, making this system more comfortable and suitable for all

amputation levels below the elbow. This is a great challenge in the EMG pattern recognition field ([3],[6], [22], [21]) since amputee patients may be discarded due to their limited available body surface.

4.5.1. Inter-session Transfer Learning

As introduced in Section 3.3.2, we applied the new networks trained on 6-electrodes inputs to day-to-day and inter-session variability using transfer learning with fine tuning. Our aim is to study the adaptability of DL algorithms, and their robustness over time and with respect to day-to-day variations. The CNNs used for transfer learning between sessions are shown in Tab. 6.

Table 6: Transfer Learning architecture. CNet_2D_2L_big (Tab. 5) is trained on the first session of one patient and the first two convolutional blocks (in grey, block 1 and block 2) are freezed. Then, the first TL-architecture (TL1) tunes both a new convolutional block (block 3) and a fully connected block with second session data. The second architecture is further simplified (TL2), directly tuning the fully-connected block without the convolutional block 3.

	Layers	**Filter size (HxW)**	**# filters**	**Output**	**Options**
1	Input			(Ch,T)	
Frozen	Conv2D	(2x15)	32	(32,Ch,T)	padding=same
	BatchNorm			(32,Ch,T)	
	RRelu			(32,Ch,T)	
	Maxpool	(1x4), stride=(1x4)		(32,Ch, T/4)	padding=valid
	Dropout			(32,Ch, T/4)	p=0.5
2	Conv2D	(2x30)	48	(48,Ch,T/4)	padding=same
Frozen	BatchNorm			(48,Ch,T/4)	
	RRelu			(48,Ch,T/4)	
	Maxpool	(1x8), stride=(1x8)		(48,Ch, T/32)	padding=valid
	Dropout			(48,Ch, T/32)	p=0.5
(3)	Input			(Ch,T)	
(TL1)	Conv2D	(2x15)	48	(48,Ch,T/32)	padding=same
	BatchNorm			(48,Ch,T/32)	
	RRelu			(48,Ch,T/32)	
	Maxpool	(1x4), stride=(1x4)		(48,Ch, T/128)	padding=valid
	Dropout			(48,Ch, T/128)	p=0.5
4	Flatten			(48xChxT/128)	
TL1	Dense		~	50	act=ReLU
TL2	BatchNorm			50	
	Dropout			50	p=0.5
	Dense			8	act=softmax

The inter-session transfer learning approach has achieved competitive results for both architectures, with 97.97% and 96.80% of accuracy, respectively. The training time required for fine-tuning was $\sim 9min$ and $\sim 4min$, respectively. The original models for the same patient trained and tested on one session achieve on average 97.80% accuracy. These results support that transfer learning is a promising approach to overcome the limitations due to inter-session variability [8, 7, 16, 10]. Furthermore, we have achieved competitive results even retraining only the dense block (TL2): this means that the automatic feature extraction stage of CNNs can be shared among sessions, supporting the superiority of DL extracted features over hand-crafted features. We reasonably believe that further improvements in accuracy can be obtained with data from more than one session for the first training phase. Employing more than one session for training, thus introducing more variability, will improve the overall performance of a transfer learning model for inter-session hand movement classification. These successful results unlock the opportunity for continuous learning for myo-controlled prostheses, not only in a laboratory-controlled condition but for real life clinical prostheses with near real-time calibrations. Indeed, with a calibration phase of less than $\sim 10\ min$ ($6-7\ min$ for data collection and $4\ min$ for tuning), it is possible to update and adapt the model to the new data. The calibration phase can be triggered daily or periodically, significantly improving the real use of the device.

4.5.2. Observing inconsistencies in electrodes' positioning

This last paragraph will describe how XAI unlocks the possibility of finding human error in a real problem. In our specific case, we have highlighted an inconsistency between XAI results and the experimental setting. The data used for this problem was collected by colleagues in the NearLab more than one year ago. The experimental setting that was firstly provided for this work had a different electrode placement. However, the global explanations obtained with generalized Grad-CAM presented in Section 4.4 highlighted that the most important channels for wrist flexion were 1,7 and 10 while for wrist extension

were 4,9,5. In the previous configuration, electrodes 1,7,10 were placed in the extensor area while electrodes 4,9,5 in the flexor muscles area. These results were not in agreement with domain knowledge, arousing questions about the correctness of the problem. Furthermore, plotting EMG signals for all channels, we found that the amplitude of the channels was again not in agreement with domain knowledge. During flexion, electrodes 1,7,10 showed larger oscillations and amplitude, suggesting that subjects were using extensor muscles for wrist flexion. Similarly, during extension, electrodes 4,5,9 were mostly activated, suggesting that the subject used flexors muscles. These observations further supported that flexion and extension seemed inverted. Further investigation, made together with the colleagues who placed the sensors, led to finding the error: the first configuration was a setting used for initial testing of the hardware while the real configuration used during the experiments was different (as shown in Fig. 2). The potential of XAI in error decoding (both human and machine) is great, and it may significantly help developers while designing a model.

5. Discussion

In this research we have applied XAI to EMG pattern recognition for hand gesture classification for myo-controlled prostheses. Explainability algorithms for EMG signal decoding have shown promising applications and we have understood the outcomes of opaque decision making systems with respect to physiological processes in the application domain of the neuromuscular system. Through our detailed explanations and visualization techniques, we have better comprehended the problem of hand gesture recognition, achieving the following insights:

I1: XAI produces physiologically highly plausible explanations of how the CNNs classify EMG. We generalized Grad-CAM to obtain global comprehensible explanations of how black-box Learning models have classified hand movements. We verified that CNNs learn to extract and use well-known EMG patterns from raw signals automatically, and they correctly identify channels

placed on muscles effectively responsible for a movement; thus, they are not relying on artifactual or random components.

I2: Hand gestures are mainly identified by higher amplitude and larger dynamic of the electrical activity of the responsible muscles. The local visualization of CNNs shows how the most important part of a signal is when the oscillations drastically increase and RMS is identified as the most important feature. Furthermore, our explanations have identified areas rather than single electrodes, supporting sensor fusions and functional synergies between muscles.

I3: The experimental set-up can be simplified by reducing the number of electrodes making the solution suitable for all under elbow amputees. In our specific setting, ten electrodes placed six proximal and four distal, are redundant. The six proximal electrodes alone are informative enough for a correct classification of eight hand gestures. Simplifying the configuration to six proximal electrodes can drastically reduce the cost of hardware, the computational time and the complexity of the system; at the same time, it minimizes the required body surface in the forearm, making this system suitable for a larger amputees population. [3, 22, 21].

I4: The features can be reduced to three meaningful features (RMS, HPM and HPC), achieving competitive performances while decreasing the inference time from 275 ms to 0.55 ms compared to the original set. We have verified the robustness and significance of the selected features and their relevance in the domain knowledge. It is of our interest, in future researches, investigating the performances, robustness and computational time of the simplified setting (with 6 electrodes) and the machine learning methods with meaningful features.

I5: Inter-session variability can be overcome by transfer learning, opening the possibility of prostheses periodically calibrated that are not prone to fail due to changes in electrodes' shift, skin conductivity or others. [8]

6. Conclusions

In this research, we applied XAI methods to opaque classifiers used for hand movement decoding, understanding how they work and gaining knowledge about the problem. As far as we know, this is the first time XAI is applied to this field. We assessed that they extract physiologically highly plausible features and characteristics well-known in domain knowledge, enhancing the **acceptance and trustworthiness** of users, domain experts, and medical professionals in AI-powered devices. Garner trust is always necessary to increase acceptance of any new technology [12, 24] and XAI could be a pivotal step to accurately assess the correctness and functioning of AI models in health-related applications. Furthermore, XAI has provided **scientific insights** about the variability of the signals, muscles' importance and EMG patterns. These insights led to three additional outputs: (1) the reduction of the number of classification features, reducing the computational complexity, leading to more robust and reliable models and paving the way for novel real-time applications, (2) the simplification of the electrode setting for better usability (controllability, dexterity and flexibility, and weight), better clinical translation (more patients), lower computational and hardware costs and (3) the detection of inconsistencies and errors in the experimental setting uncovered comparing the algorithms explanations with the presumed configuration of the EMG sensors. All these achievements have been possible only through the cooperation between artificial intelligence and human intelligence. Human knowledge and judgment have been fundamental in analyzing and validating the results, in error debugging, and in drawing scientific insights from mathematical results. *"Human in the loop"* is a key concept for the future of AI and its application in critical fields such as medicine and neuroscience, where experts' opinion is valuable and deep-rooted.

References

[1] P. Xia, J. Hu, Y. Peng, Emg-based estimation of limb movement using deep learning with recurrent convolutional neural networks, Artificial organs 42 (5) (2018) E67–E77.

[2] M. Zia ur Rehman, A. Waris, S. O. Gilani, M. Jochumsen, I. K. Niazi, M. Jamil, D. Farina, E. N. Kamavuako, Multiday emg-based classification of hand motions with deep learning techniques, Sensors 18 (8) (2018) 2497.

[3] A. H. Al-Timemy, G. Bugmann, J. Escudero, N. Outram, Classification of finger movements for the dexterous hand prosthesis control with surface electromyography, IEEE Journal of Biomedical and Health Informatics 17 (3) (2013) 608–618.

[4] A. Phinyomark, P. Phukpattaranont, C. Limsakul, Feature reduction and selection for emg signal classification, Expert Systems with Applications 39 (8) (2012) 7420 – 7431. doi:https://doi.org/10.1016/j.eswa.2012.01.102.
URL http://www.sciencedirect.com/science/article/pii/S0957417412001200

[5] M. Atzori, A. Gijsberts, C. Castellini, B. Caputo, A.-G. Mittaz Hager, S. Elsig, G. Giatsidis, F. Bassetto, H. Müller, Electromyography data for non-invasive naturally-controlled robotic hand prostheses, Nature 1 (12 2014). doi:10.1038/sdata.2014.53.

[6] M. Zia ur Rehman, S. O. Gilani, A. Waris, I. K. Niazi, E. N. Kamavuako, A novel approach for classification of hand movements using surface emg signals, in: 2017 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), 2017, pp. 265–269.

[7] O. W. Samuel, H. Zhou, X. Li, H. Wang, H. Zhang, A. K. Sangaiah, G. Li, Pattern recognition of electromyography signals based

on novel time domain features for amputees' limb motion classification, Computers Electrical Engineering 67 (2018) 646–655. doi:https://doi.org/10.1016/j.compeleceng.2017.04.003.
URL https://www.sciencedirect.com/science/article/pii/S0045790617303932

[8] O. Samuel, M. Asogbon, Y. Geng, A. Al-Timemy, S. Pirbhulal, N. Ji, S. Chen, P. Fang, P. Li, Intelligent emg pattern recognition control method for upper-limb multifunctional prostheses: Advances, current challenges, and future prospects, IEEE Access PP (2019) 1–1. doi:10.1109/ACCESS.2019.2891350.

[9] K. Park, S. Lee, Movement intention decoding based on deep learning for multiuser myoelectric interfaces, in: 2016 4th International Winter Conference on Brain-Computer Interface (BCI), 2016, pp. 1–2.

[10] Y. Du, W. Jin, W. Wei, Y. Hu, W. Geng, Surface emg-based inter-session gesture recognition enhanced by deep domain adaptation, Sensors (Basel, Switzerland) 17 (2017).

[11] W. Li, P. Shi, H. Yu, Gesture recognition using surface electromyography and deep learning for prostheses hand: State-of-the-art, challenges, and future, Frontiers in Neuroscience 15 (2021) 259. doi:10.3389/fnins.2021.621885.
URL https://www.frontiersin.org/article/10.3389/fnins.2021.621885

[12] A. Barredo Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila, F. Herrera, Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai, Information Fusion 58 (2020) 82 – 115. doi:https://doi.org/10.1016/j.inffus.2019.12.012.

URL http://www.sciencedirect.com/science/article/pii/S1566253519308103

[13] N. Burkart, M. F. Huber, A survey on the explainability of supervised machine learning, Journal of Artificial Intelligence Research 70 (2021) 245–317. doi:10.1613/jair.1.12228.
URL http://dx.doi.org/10.1613/jair.1.12228

[14] A. Phinyomark, F. Quaine, S. Charbonnier, C. Serviere, F. Tarpin-Bernard, Y. Laurillau, Emg feature evaluation for improving myoelectric pattern recognition robustness, Expert Systems with Applications 40 (12) (2013) 4832 – 4840. doi:https://doi.org/10.1016/j.eswa.2013.02.023.
URL http://www.sciencedirect.com/science/article/pii/S0957417413001395

[15] H.-m. Shim, H. An, S. Lee, E. Lee, H.-k. Min, S. Lee, Emg pattern classification by split and merge deep belief network, Symmetry 8 (2016) 148. doi:10.3390/sym8120148.

[16] M. Zia ur rehman, M. Waris, S. Gilani, M. Jochumsen, I. Niazi, M. Jamil, D. Farina, E. Kamavuako, Multiday emg-based classification of hand motions with deep learning techniques, Sensors 18 (08 2018). doi:10.3390/s18082497.

[17] U. Côté-Allard, E. Campbell, A. Phinyomark, F. Laviolette, B. Gosselin, E. Scheme, Interpreting deep learning features for myoelectric control: A comparison with handcrafted features, Frontiers in bioengineering and biotechnology 8 (2020) 158.

[18] M. Tavakoli, C. Benussi, J. L. Lourenco, Single channel surface emg control of advanced prosthetic hands: A simple, low cost and efficient approach, Expert Systems with Applications 79 (2017) 322–332.

[19] N. Parajuli, N. Sreenivasan, P. Bifulco, M. Cesarelli, S. Savino, V. Niola, D. Esposito, T. J. Hamilton, G. R. Naik, U. Gunawardana, et al., Real-

time emg based pattern recognition control for hand prostheses: a review on existing methods, challenges and future implementation, Sensors 19 (20) (2019) 4596.

[20] F. V. Tenore, A. Ramos, A. Fahmy, S. Acharya, R. Etienne-Cummings, N. V. Thakor, Decoding of individuated finger movements using surface electromyography, IEEE transactions on biomedical engineering 56 (5) (2008) 1427–1434.

[21] H. Huang, P. Zhou, G. Li, T. A. Kuiken, An analysis of emg electrode configuration for targeted muscle reinnervation based neural machine interface, IEEE Transactions on Neural Systems and Rehabilitation Engineering 16 (1) (2008) 37–45. doi:10.1109/TNSRE.2007.910282.

[22] G. R. Naik, A. H. Al-Timemy, H. T. Nguyen, Transradial amputee gesture classification using an optimal number of semg sensors: An approach using ica clustering, IEEE Transactions on Neural Systems and Rehabilitation Engineering 24 (8) (2016) 837–846. doi:10.1109/TNSRE.2015.2478138.

[23] U. Côté-Allard, C. L. Fall, A. Drouin, A. Campeau-Lecours, C. Gosselin, K. Glette, F. Laviolette, B. Gosselin, Deep learning for electromyographic hand gesture signal classification using transfer learning, IEEE Transactions on Neural Systems and Rehabilitation Engineering 27 (4) (2019) 760–771.

[24] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, D. Pedreschi, A survey of methods for explaining black box models, ACM Comput. Surv. 51 (5) (Aug. 2018). doi:10.1145/3236009.
URL https://doi.org/10.1145/3236009

[25] A. Holzinger, G. Langs, H. Denk, K. Zatloukal, H. Müller, Causability and explainability of artificial intelligence in medicine, WIREs Data Mining and Knowledge Discovery 9 (4) (2019) e1312. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/widm.1312, doi:10.1002/widm.1312.
URL https://onlinelibrary.wiley.com/doi/abs/10.1002/widm.1312

[26] M. T. Ribeiro, S. Singh, C. Guestrin, "why should I trust you?": Explaining the predictions of any classifier, CoRR abs/1602.04938 (2016). arXiv:1602.04938.
URL http://arxiv.org/abs/1602.04938

[27] S. M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, in: I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett (Eds.), Advances in Neural Information Processing Systems 30, Curran Associates, Inc., 2017, pp. 4765–4774.

[28] A. Shrikumar, P. Greenside, A. Kundaje, Learning important features through propagating activation differences, CoRR abs/1704.02685 (2017). arXiv:1704.02685.
URL http://arxiv.org/abs/1704.02685

[29] R. R. Selvaraju, A. Das, R. Vedantam, M. Cogswell, D. Parikh, D. Batra, Grad-cam: Why did you say that? visual explanations from deep networks via gradient-based localization, CoRR abs/1610.02391 (2016). arXiv:1610.02391.
URL http://arxiv.org/abs/1610.02391

[30] S. Woo, J. Park, J.-Y. Lee, I. S. Kweon, Cbam: Convolutional block attention module, in: Proceedings of the European conference on computer vision (ECCV), 2018, pp. 3–19.

[31] I. Sturm, S. Lapuschkin, W. Samek, K.-R. Müller, Interpretable deep neural networks for single-trial eeg classification, Journal of Neuroscience Methods 274 (2016) 141 – 145. doi:https://doi.org/10.1016/j.jneumeth.2016.10.008.
URL http://www.sciencedirect.com/science/article/pii/S0165027016302333

[32] R. T. Schirrmeister, J. T. Springenberg, L. D. J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard,

T. Ball, Deep learning with convolutional neural networks for eeg decoding and visualization, Human Brain Mapping 38 (11) (2017) 5391–5420. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/hbm.23730, doi:10.1002/hbm.23730.
URL https://onlinelibrary.wiley.com/doi/abs/10.1002/hbm.23730

[33] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, B. J. Lance, Eegnet: A compact convolutional network for eeg-based brain-computer interfaces, CoRR abs/1611.08024 (2016). arXiv:1611.08024.
URL http://arxiv.org/abs/1611.08024

[34] S. Sakhavi, C. Guan, S. Yan, Learning temporal information for brain-computer interface using convolutional neural networks, IEEE Transactions on Neural Networks and Learning Systems 29 (11) (2018) 5619–5629.

[35] K. G. Hartmann, R. T. Schirrmeister, T. Ball, Hierarchical internal representation of spectral features in deep convolutional networks trained for EEG decoding, CoRR abs/1711.07792 (2017). arXiv:1711.07792.
URL http://arxiv.org/abs/1711.07792

[36] Y. Yu, C. Chen, J. Zhao, X. Sheng, X. Zhu, Surface electromyography image driven torque estimation of multi-dof wrist movements, IEEE Transactions on Industrial Electronics (2021).

[37] R. Soroushmojdehi, S. Javadzadeh, Classification of emg signals for hand movement intention detection, Master thesis, Politecnico di Milano, Milano (2020).

[38] D. Stegeman, H. Hermens, Standards for surface electromyography: The european project surface emg for non-invasive assessment of muscles (seniam), Enschede: Roessingh Research and Development (2007) 108–12.

[39] J. Wang, L. Tang, J. E. Bronlund, Surface emg signal amplification and filtering, International Journal of Computer Applications 82 (1) (2013).

[40] B. Hudgins, P. Parker, R. N. Scott, A new strategy for multifunction myoelectric control, IEEE Transactions on Biomedical Engineering 40 (1) (1993) 82–94.

[41] K. Englehart, B. Hudgins, P. Parker, M. Stevenson, Classification of the myoelectric signal using time-frequency based representations, Medical Engineering & Physics 21 (6) (1999) 431 – 438. doi:https://doi.org/10.1016/S1350-4533(99)00066-1.
URL http://www.sciencedirect.com/science/article/pii/S1350453399000661

[42] M. Mouzé-Amady, F. Horwat, Evaluation of hjorth parameters in forearm surface emg analysis during an occupational repetitive task, Electroencephalography and Clinical Neurophysiology/Electromyography and Motor Control 101 (2) (1996) 181–183.

[43] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (2015) 436–44. doi:10.1038/nature14539.

[44] C. Yang, A. Rangarajan, S. Ranka, Global model interpretation via recursive partitioning, in: 2018 IEEE 20th International Conference on High Performance Computing and Communications; IEEE 16th International Conference on Smart City; IEEE 4th International Conference on Data Science and Systems (HPCC/SmartCity/DSS), IEEE, 2018, pp. 1563–1570.

[45] N. Frosst, G. Hinton, Distilling a neural network into a soft decision tree, arXiv preprint arXiv:1711.09784 (2017).

[46] D. Janzing, L. Minorics, P. Blöbaum, Feature relevance quantification in explainable ai: A causal problem, in: International Conference on Artificial Intelligence and Statistics, PMLR, 2020, pp. 2907–2916.

[47] A. Mahendran, A. Vedaldi, Salient deconvolutional networks, in: European Conference on Computer Vision, Springer, 2016, pp. 120–135.

[48] H. Fukui, T. Hirakawa, T. Yamashita, H. Fujiyoshi, Attention branch network: Learning of attention mechanism for visual explanation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 10705–10714.

[49] Y. Zhang, D. Hong, D. McClement, O. Oladosu, G. Pridham, G. Slaney, Grad-cam helps interpret the deep learning models trained to classify multiple sclerosis types using clinical brain magnetic resonance imaging, Journal of Neuroscience Methods 353 (2021) 109098.

[50] S. Jonas, A. O. Rossetti, M. Oddo, S. Jenni, P. Favaro, F. Zubler, Eeg-based outcome prediction after cardiac arrest with convolutional neural networks: Performance and visualization of discriminative features, Human brain mapping 40 (16) (2019) 4606–4617.

[51] Y. Li, H. Yang, J. Li, D. Chen, M. Du, Eeg-based intention recognition with deep recurrent-convolution neural network: Performance and channel selection by grad-cam, Neurocomputing 415 (2020) 225–233.

[52] C. Molnar, Interpretable machine learning, in: Interpretable Machine Learning:A Guide for Making Black Box Models Explainable, 2019, https://christophm.github.io/interpretable-ml-book/.

[53] R. Boostani, M. Moradi, Evaluation of the forearm emg signal features for the control of a prosthetic hand, Physiological measurement 24 (2003) 309–19. doi:10.1088/0967-3334/24/2/307.

[54] M. Zardoshti-Kermani, B. C. Wheeler, K. Badie, R. M. Hashemi, Emg feature evaluation for movement control of upper extremity prostheses, IEEE Transactions on Rehabilitation Engineering 3 (4) (1995) 324–333. doi:10.1109/86.481972.

[55] V. Losing, B. Hammer, H. Wersing, Incremental on-line learning: A review and comparison of state of the art algorithms, Neurocomputing 275 (09 2017). doi:10.1016/j.neucom.2017.06.084.

[56] Y. Liu, J. Wang, J. Li, S. Niu, H. Song, Class-incremental learning for wireless device identification in iot (2021). arXiv:2105.06381.

[57] M. Asghari Oskoei, H. Hu, Myoelectric control systems—a survey, Biomedical Signal Processing and Control 2 (4) (2007) 275 – 294. doi:https://doi.org/10.1016/j.bspc.2007.07.009.
URL http://www.sciencedirect.com/science/article/pii/S1746809407000547

[58] M. A. Ozdemir, G. D. Ozdemir, O. Guren, Classification of covid-19 electrocardiograms by using hexaxial feature mapping and deep learning, BMC Medical Informatics and Decision Making 21 (1) (2021) 1–20.

[59] M. A. Ozdemir, M. Degirmenci, E. Izci, A. Akan, Eeg-based emotion recognition with deep convolutional neural networks, Biomedical Engineering/Biomedizinische Technik 66 (1) (2021) 43–57.

[60] S. M. Lundberg, G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, S.-I. Lee, From local explanations to global understanding with explainable ai for trees, Nature Machine Intelligence 2 (1) (2020) 2522–5839.

[61] M. Verleysen, D. François, The curse of dimensionality in data mining and time series prediction, Vol. 3512, 2005, pp. 758–770. doi:10.1007/11494669_93.

Appendix A. Features

This section presents the features selected in this work. Unless specified otherwise, features are calculated by dividing the signal x into overlapping windows of length L. The k^{th} element of the i^{th} window then corresponds to $x_{i,k}$.

- Mean Absolute Value

$$MAV(x_i) = \frac{1}{L}\sum_{k=1}^{L}|x_{i,k}| \tag{A.1}$$

- Zero Crossing, number of times the signal crosses the zero
- Slope Sign Change, number of times the sign of the signal changes
- Waveform Length

$$WL(x_i) = \sum_{k=1}^{L}|x_{i,k} - x_{i,k-1}| \tag{A.2}$$

- Hjorth Parameter - Activity

$$A(x_i) = \frac{1}{L}\sum_{k=1}^{L}(x_{i,k} - \overline{x}_i)^2 \tag{A.3}$$

- Hjorth Parameter - Mobility

$$M(x_i) = \sqrt{\frac{A(x_i^{'})}{A(x_i)}} \tag{A.4}$$

  with x^i the firt derivative with respect to time.

- Hjorth Parameter - Complexity

$$C(x_i) = \frac{M(x_i^{'}}{x_i} \tag{A.5}$$

- Sample Entropy

$$SampEn(x_i, m, r) = -\ln\frac{A^m(r)}{B^m(r)} \tag{A.6}$$

  with m the embedding dimension and r the tolerance.

- Cepstral coefficients: they can be derived from AR coefficients as follows:

$$c_1 = -a_1 \tag{A.7}$$

$$c_i = a_i - \sum_{n=1}^{i-1}(1 - \frac{n}{i}) * a_n * c_{i-n} \tag{A.8}$$

with $1 < i \leq P$, C_i is the i^{th} Cepstral coefficient, a_i is the i^{th} auto-regression coefficient and P is the order.

- Root Mean Square

$$RMS(x_i) = \sqrt{\frac{1}{L}\sum_{k=1}^{L} x_{i,k}^2} \tag{A.9}$$

- Integrated EMG

$$IEMG(x_i) = \sum_{k=1}^{L} |x_{i,k}| \tag{A.10}$$

- Skewness

$$Skew(x_i) = \frac{1}{L}\sum_{k=1}^{L}(\frac{x_{i,k} - \overline{x}_i}{\sigma})^3 \tag{A.11}$$

Appendix B. ML hyperparameters and results

Table B.7: Classical Machine Learning Hyperparameters. Hyperparameter tuning has been performed with a 3-cv cross validation on the training set on a randomly chosen subject. The best set was selected based on average accuracy.

Classifier	Parameter 1	Parameter 2	Parameter 3	Parameter 4
KNN	K: 10, 20, 30, 40, 50	weights: uniform distance		
MLP	hidden layers: (100,50,20), (50,20), (10,8), (100), (200), (100,20)	alpha: 0.001, 0.0001, 0.00001	activation function: tanh, identity, relu, logistic	solver: sgd, Adam
SVM	regularization: 0.1, 1, 10, 100	kernel: linear, poly, rbf, sigmoid	degree for poly: 1,2,3,4	gamma: auto, 0.1, 10e-7
LDA	solver: svd, lsqr, eigen			
XRT	number estimators: 10, 50, 100, 150	criterion gini, entropy		
Bagging	number estimators: 10, 50, 100, 150	Base estimator Decision Tree Classifier (max_depth=10, 15, 30)		
Adaboost	number estimators: 10, 50, 100, 150	Base Estimator Decision Tree Classifier (max_depth=10, 15, 30)		
Random Forest	number estimators: 10, 50, 100, 150	criterion gini, entropy		

Table B.8: Classical Machine Learning results in term of average Precision (%), Recall (%) and inter-subject standard deviation

	Precision		Recall	
Classifier	Improved TD	FULL	Improved TD	FULL
LDA	92.23 ±3.67	93.11 ± 3.64	91.16 ±4.28	92.10 ± 4.48
SVM	92.38 ± 3.12	92.10 ±3.08	91.03 ± 3.69	90.79 ±3.35
XRT	90.66 ± 3.51	90.60 ±3.63	89.60 ± 3.85	89.27 ±3.79

Table B.9: Classical Machine Learning results of all models in term of average accuracy (%), f1 score (%) on the two sets of data employed (Improved TD and FULL)

	Accuracy		F1-score	
Classifier	Improved TD	FULL	Improved TD	FULL
KNN	89.20 ±4.35	88.70 ± 4.67	89.26 ± 4.12	88.63 ± 4.47
MLP	91.45 ± 2.97	91.26 ±3.15	91.32 ±2.96	91.23 ±3.22
Bagging	85.91 ± 3.49	85.77 ± 3.81	85.95 ± 3.54	85.64 ± 3.90
Random Forest	88.75 ± 3.85	88.95 ± 3.95	88.65 ± 3.43	88.84 ± 3.90
Adaboost	84.66 ± 2.70	83.73 ± 2.93	84.17 ± 3.22	83.49 ± 3.10

Appendix C. CNET architectures

Table C.10: CNet architecture, with H=1 for CNet1D and H=2 for CNet2D. CNet1D applies filtering only in the time dimension (along the width of the input) while CNet2D performs a 2D filtering both in time (width) and space (heigth).

	Layers	**Filter size (HxW)**	**# filters**	**Output**	**Options**
1	Input			(Ch,T)	
	Conv	(Hx13)	32	(32,Ch,T)	padding=same
	BatchNorm			(32,Ch,T)	
	RRelu			(32,Ch,T)	
	Maxpool	(1x4), stride=(1x4)		(32,Ch, T/4)	padding=valid
	Dropout			(32,Ch, T/4)	p=0.5
2	Conv	(Hx9)	48	(48,Ch,T/4)	padding=same
	BatchNorm			(48,Ch,T/4)	
	RRelu			(48,Ch,T/4)	
	Maxpool	(1x4), stride=(1x4)		(48,Ch, T/16)	padding=valid
	Dropout			(48,Ch, T/16)	p=0.5
3	Conv	(Hx5)	64	(64,Ch,T/16)	padding=same
	BatchNorm			(64,Ch,T/16)	
	RRelu			(64,Ch,T/16)	
	Maxpool	(1x4), stride=(1x4)		(64,Ch, T/64)	padding=valid
	Dropout			(64,Ch, T/64)	p=0.5
4	Flatten			(64xChxT/64)	
	Dense			300	
	BatchNorm			300	
	RRelu			300	
	Dropout			300	p=0.5
	Dense			50	
	BatchNorm			50	
	RRelu			50	
	Dropout			50	p=0.5
	Dense			8	activation=softmax