Framework for Learning a Hand Intent Recognition Model from sEMG for FES-based control

Neha Das[1], Satoshi Endo[1], Hossein Kavianirad[1] and Sandra Hirche[1]

***Abstract*— Stroke survivors and individuals with neuromuscular disorders often experience motor function impairments, particularly during hand movements crucial for activities of daily living (ADL). Functional Electrical Stimulation (FES) has emerged as a potential assistive and rehabilitative technique to address these limitations. However, accurately determining user intent during FES poses a significant challenge. This work proposes a framework for rapidly learning a model of the user's hand intent from surface electromyography (sEMG) signals, specifically for continuous FES-based control of the ipsilateral hand. The framework systematically collects data from expected volitional and FES-evoked hand motions, followed by training a logistic regression model for intent classification. The study demonstrates that the proposed model can learn from limited data and compares favorably to deep neural nets trained on the same dataset. This model is able to recognize user intent with high accuracy even during concurrent FES stimulation.**

I. Introduction

Stroke survivors, as well as patients of other neuromuscular disorders may suffer from motor function impairments that can severely limit their capability to perform activities of daily living (ADL) [1]. In particular, these impairments include the ability to close and open one's hand [2] - a functionality that is integral to many ADL tasks, such as eating, drinking and dressing which all require dexterous manipulation of objects. Functional electrical stimulation (FES) has emerged as a potential assistive and rehabilitative technique [3]. FES is a technique that can artificially activate muscles by providing small electrical impulses through the user's skin [4]. Consequently, relevant motions can be generated in the user's hand with FES activation. However, to facilitate closed-loop FES control, a feedback mechanism is required to inform the controller about the deviation from the desired system state. For assistive control, therefore, the user's intention needs to be estimated.

*This work was supported by the European Research Council (ERC) Consolidator Grant "Safe data-driven control for human-centric systems (CO-MAN)" under grant agreement number 864686, by the Horizon 2020 research and innovation programme of the European Union under grant agreement number 871767 of the project ReHyb: Rehabilitation based on hybrid neuroprosthesis, and by TUM AGENDA 2030, funded by the Federal Ministry of Education and Research (BMBF) and the Free State of Bavaria under the Excellence Strategy of the Federal Government and the Länder as well as by the Hightech Agenda Bavaria.

[1]Chair of Information-oriented Control, School of Computation, Information and Technology, Technical University of Munich, 80992 Munich, Germany. E-mails: {neha.das, s.endo, hossein.kavianirad, hirche}@tum.de

Article accepted for BioRob 2024 conference proceedings

Previous works have exploited many sensors and techniques for inferring the user's intent. These include classification of hand intent from optical marker-based motion tracking [5], [6], video data [7], [8], pose-estimation via inertial sensors [9] and muscle activity measurement via surface electromyography (sEMG) [10]–[14]. However, the evoked hand pose via FES stimulation might not be distinguishable from hand poses obtained by volitional motion since they result in similar kinematic motions. Thus volitional intent cannot be reliably extracted during FES stimulation from these methods.

Identification of volitional intent during concurrent FES stimulation can be addressed by methods such as Brain-Computer Interface (BCI) [15], [16], as well as explicit communication of intent via the motion of the non-paretic limb [17]. Compared to passive and triggered FES control, these methods have been known to gain relatively higher motor function improvement. This indicates that active participation by the user in the FES control loop, such that the FES behavior is aligned with the motor intention may be a key ingredient for successful FES based rehabilitation. However, compared to BCI and contralateral control which involve either estimating intent from cognition or explicitly obtaining the user's intent, sEMG offers a comparatively direct measure of this intention as it forms the final evidence for attempted movement in the form of muscle excitation [18]. Considering all the above factors, active FES control of the impaired limb based on measurements from ipsilaterally placed sEMG electrodes can provide an adequate solution if the volitional sEMG signal can be successfully extracted.

Previous studies have investigated several methods for determining hand-intent from sEMG signals. Heuristic methods that classify motion intent by thresholding the value of sEMG signal from distinct channels have been used by [11]. However, these methods rely on thresholds calculated from a certain population that need to be adjusted for new users. Various data-driven techniques have also been explored, including feature-based classifiers such as those presented by [10], [12], [19]. These approaches extract characteristic pre-defined features from the sEMG signals and combine them linearly or non-linearly with learnable parameters and a link function to predict the motion intent class. Frequently used features encompass time-domain characteristics [20], [21] such as the signal's mean absolute value, root mean square, slope sign change, and waveform length. Frequency domain features [22] such as power spectral density, median-frequency, auto-regression coefficients, as well as time-frequency features [23], such as wavelet transforms have

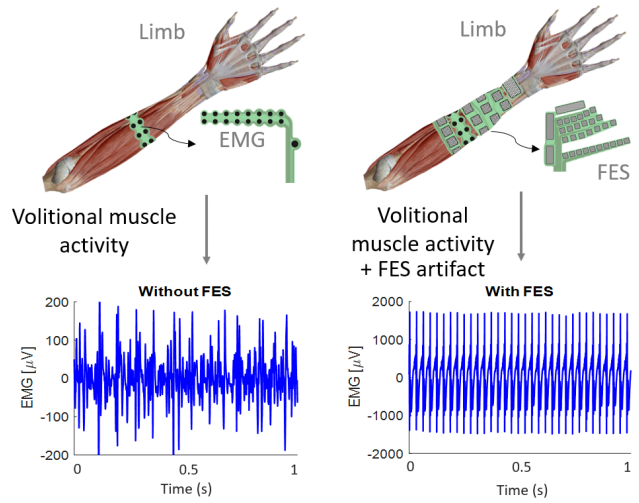


Fig. 1: Illustration of the sEMG signal corruption with FES artifacts when FES is stimulating the same muscle group whose activity the sEMG device is recording.

also been employed. Since the calculation of frequency domain features is computationally more demanding, on-line performance in classification often prioritizes time-domain features [20].

In contrast to fixed feature based approaches, deep-learning methods [24], [25] deploy specialized deep networks such as convolutional neural networks (CNN) and fully connected networks (FCN) to automatically extract the features from the signal, thus surpassing the need for feature engineering. However, a potential drawback of deep-learning techniques is their reliance on large datasets for accurate model learning; inadequate data may hinder their generalization. Therefore, quick calibration of an intent recognition model from a small data set may be difficult for these approaches.

Many of the above approaches for classifying hand intent from sEMG do not address scenarios where sEMG signals are influenced by artifacts generated through FES. This can be a common occurrence during FES control of a user's hand using sEMG sensors, where FES may concurrently be activated during sEMG recording. FES stimulation typically alters the sEMG measurements by introducing artifacts to the signal that can dominate the volitional sEMG signal [26]. Figure 1 illustrates this effect. These artifacts include stationary and short-lived effects such as the transient stimulation and non-stationary effects such as M-wave that can persist throughout the recording process. Consequently, an intent classifier learned from volitional sEMG data may misconstrue the user's intent when processing an sEMG signal corrupted due to FES stimulation.

To address this issue, various approaches have been explored to extract intentional sEMG from such signals, including hardware methods such as blanking windows [27] and software methods such as filtering [12], [26], [28] and decomposition [29]. Of these, adaptive filtering techniques, such as those proposed by [26] and [28], as well as an iteration of such algorithms based on Gram-Schmidt algorithm [28], [30] have been widely adopted since they offer an online approach for mitigating both stationary and non-stationary artifacts. In contrast, methods such as blanking and fixed-response comb filters may either lead to data loss or struggle to completely eliminate non-stationary artifacts introduced by FES stimulation or both.

Even with the adaptive filter applied, however, we observed that signals varied depending on the location and intensity of stimulation for the same volitional motion. This indicates that the adaptive filter is unable to remove all the effects of FES stimulation from the recorded sEMG signal. Consequently, an intent recognition model learned solely from volition sEMG recordings may give inaccurate results when tested on sEMG signals evoked from both volitional motion and FES, even if it has been filtered to mitigate FES-produced artifacts.

Therefore, in this work, we propose a framework for learning a model of a user's hand intent from sEMG collected from a forearm for continuous FES-based control of their ipsilateral hand. Three hand-motion categories are considered in this approach - *hand relax*, where the hand is at a rest state, *hand close*, where the fingers are flexed to closed the hand, and *hand open*, where the fingers are extended to open the hand. These three categories are chosen since hand opening and closing form an important submodule of ADL tasks [2], especially tasks which require handling and manipulation of objects. The goal of our framework is to learn an intent recognition model that is robust in the presence of FES simulations evoking different motions. This is ensured by systematically collecting different combinations of expected volitional and FES-evoked hand motions for a short time period, followed by learning an intent classifier from this data using a logistic regression model. We demonstrate that a model learned from limited amounts of sEMG signals acquired with our framework yields higher performance for recognizing user intent (relax, close or open) in the presence of FES, when compared to deep neural nets (CNN and FCN) and models that exclude FES-evoked hand motions from their training set.

II. Materials and Methods

A. Experimental Setup

Ten healthy participants without any history of neurological disorders were recruited for the study. Table I summarizes the demographic information for all the participants. All participants used their right arm for the study.

The experimental setup consisted of a bluetooth-enabled sEMG device (Tecnalia Research & Innovation, Spain) with an array of 8 bipolar electrodes that can measure potentials arising from muscular activity non-invasively through the surface of the user's skin. The electrode array for the sEMG device was connected to an acquisition module containing an analog to digital converter for biopotential measurements. The device has a recording frequency of 1000 Hz. The sEMG bipolar electrode array is shown in Figure 2(b). An FES device (Tecnalia Research & Innovation, Spain) with

TABLE I: Demographic information of the participants

Characteristics	Distribution (Mean ± Std.)
Number of participants	10 (9 Male, 1 Female)
Age (years)	26.50 ± 1.02
Weight (kg)	81.80 ± 11.24
Height (cm)	182.90 ± 8.23

a 32-electrode array consisting 29 cathodes and 3 anodes was also included in the experimental setup. The electrode array for the FES device was connected to a bluetooth-enabled stimulation device that allowed for the activation and deactivation of the electrode pad and online manipulation of stimulation intensity. The FES electrode array is shown in Figure 2 (a). The stimulation frequency for all participants was kept at 25 Hz.

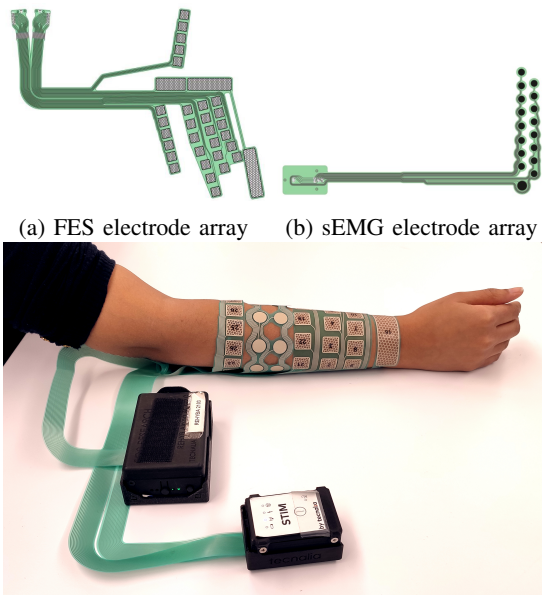
Before the placement of the sEMG and FES electrodes, the skin on the participant's forearm was cleaned with a damp cloth. The ulnar styloid apophysis was located on the participant's wrist as a reference for the placement of the FES electrodes and was used to place the wrist anode part of the FES electrode array horizontally along the wrist. The anodes were placed along the ulna next to the first anode. The rest of the FES electrode array was then wrapped around the participant's forearm. The width of the FES array and its electrode density ensured that with proper initial placement of the array along the length of the forearm, stimulating both hand close and open motions were possible. The sEMG electrode array was placed on the forearm for measuring volitional and stimulated muscle activations corresponding to different hand motions. The placement of both FES and sEMG electrode arrays is shown in Figure 2(c).

One of the pre-requisites of our framework is that the stimulation sites, i.e., the electrode pads in the FES array to be activated, and stimulation intensity for evoking each motion primitive is determined via manual calibration for each participant separately. This calibration was therefore performed before data collection. The best stimulation sites for evoking a motion were found through manual search. The stimulation amplitude was increased incrementally until the appropriate motion was evoked or the participant experienced discomfort. The amplitude just before the participant started experiencing discomfort was used.

B. Signal pre-processing via adaptive filter

FES stimulation can cause several artifacts to develop in the recorded sEMG signal. These include:

- the transient stimulation, which is a direct effect of the FES pulse that results in a large initial spike in the observed sEMG voltage. This effect lasts for about 2-3 milliseconds before exponentially decaying,
- the conduction latency, which depicts the dissipation time of initial spike and varies with the distance between the detection site and the innervation zone, and
- the M-wave, which results from the simultaneous activation of all motor units due to innervation by the FES pulse. This effect is typically non-stationary and can persist throughout the signal for high stimulation frequencies.



(a) FES electrode array (b) sEMG electrode array

(c) Placement of the electrode arrays on participant forearm

Fig. 2: **Experimental Setup:** (a) and (b) show the FES and sEMG electrode array respectively. (c) shows the placement of the electrode arrays on the right forearm of the participant.

These artifacts can dominate the sEMG signal, thereby obfuscating any present volitional sEMG. In order to effectively identify and use the volitional components of the affected signal for intent recognition, these artifacts must be alleviated. To this end, we employ adaptive filtering based on [26] and used in [31], [32] to filter the affected sEMG. This particular approach was chosen since it has been shown to effectively mitigate non-stationary effects such as the M-wave as compared to simpler filtering schemes, removes the necessity of hardware-based interventions, and is computationally more efficient than decomposition-based approaches.

The adaptive filter relies on a weighted linear combination of previous frames of the sEMG signal to predict the FES-induced artifacts in the current frame. The weights of the filter are determined through an optimization problem that minimizes the energy of the resultant filtered signal. We refer the reader to [32] for details regarding the calculation of this filter. Figure 3 illustrates an example of adaptive filtering for extracting the volitional signal from recorded sEMG during FES application.

C. Feature extraction for classification

Following the application of the adaptive filter, a set of six time-domain features characterizing the filtered signal are calculated. Since the hand intent-recognition algorithm is meant to be used in a continuous control setting, these features are calculated over sliding windows of length K of

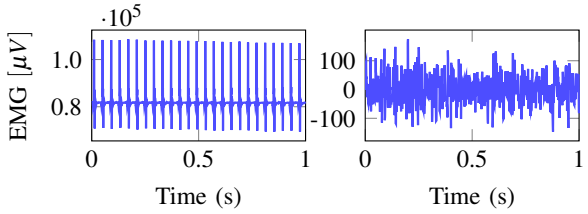


Fig. 3: Adaptive Filtering: The left column shows 1 second of an sEMG signal segment corresponding to volitional motion by the user to *open* their hand while FES stimulation for evoking the same motion is applied to them. The right column shows the filtered signal after applying the adaptive filtering scheme.

the recorded signal. Thus, at test time, feature extraction can be performed over a K length buffer that is updated with a given frequency as the signal from the sEMG device is streamed. These features include:

1) Mean absolute value (MAV): A moving average of the absolute value of the signal. For the t-th segment $\boldsymbol{x}_t$ of the signal, MAV can be formulated as follows.

$$MAV(t) = \frac{1}{K}\sum_{k=1}^{K}|x_{t,k}| \tag{1}$$

where $x_{t,k}$ represents the k-th sample in the segment.

2) Mean absolute value slope (MAVS): The difference between the MAV of the current and the previous segment. MAVS can be given by the following equation.

$$MAVS(t) = MAV(t) - MAV(t-1) \tag{2}$$

3) Zero-crossings (ZC): The number of times the waveform changes sign. This feature may capture some of the frequency characteristics of the signal, since larger number of zero-crossings imply large frequencies. ZC can be given by the following equation.

$$ZC(t) = \sum_{k=2}^{K}\left[\begin{cases}1 & \text{if } (x_{t,k-1} > 0 \text{ and } x_{t,k} < 0) \text{ or} \\ & (x_{t,k-1} < 0 \text{ and } x_{t,k} > 0) \\ 0 & \text{otherwise}\end{cases}\right] \tag{3}$$

4) Slope sign change (SSC): The number of times the slope of the waveform changes sign. Similar to ZC, this feature may also capture some of the frequency characteristics of the signal. SSC can be formulated as follows:

$$\begin{aligned}\Delta\boldsymbol{x}_t &= [x_{t,k} - x_{t,k-1}]_{k=2}^{K} \\ SSC(t) &= \sum_{k'=2}^{K-1}\left[\begin{cases}1 & \text{if } (\Delta x_{t,k'-1} > 0 \text{ and } \Delta x_{t,k'} < 0) \text{ or} \\ & (\Delta x_{t,k'-1} < 0 \text{ and } \Delta x_{t,k'} > 0) \\ 0 & \text{otherwise}\end{cases}\right]\end{aligned} \tag{4}$$

5) Waveform length (WL): The cumulative length of the waveform over the t-th segment. This feature attempts to encapsulate the complexity of the waveform, that is, its frequency and amplitude in a single measure [20]. WL can be depicted by the following equation.

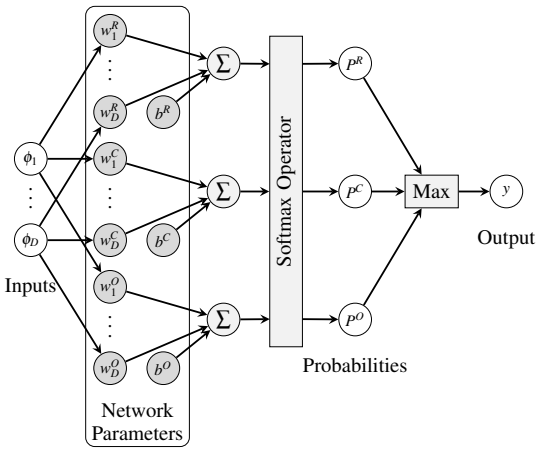


Fig. 4: Architecture of the multinomial logistic regressor. The F features of an input are linearly combined with 3 sets of parameters denoted by superscripts R, C and O, respectively corresponding to hand relax, hand close and hand open.

$$\begin{aligned}\Delta\boldsymbol{x}_t &= [x_{t,k} - x_{t,k-1}]_{k=2}^{K} \\ WL(t) &= \sum_{k'=1}^{K-1}|\Delta x_{t,k'}|\end{aligned} \tag{5}$$

6) Root mean square (RMS): The root mean square of the signal amplitude over the t-th segment. The RMS can also be seen as a function of the energy of the signal segment [19]. RMS can be given as follows.

$$RMS(t) = \sqrt{\frac{1}{K}\sum_{k=1}^{K}x_{t,k}^2} \tag{6}$$

A subset of the above features can be calculated from a signal segment $\boldsymbol{x}_t$ and concatenated in a vector $\boldsymbol{\phi}(\boldsymbol{x}_t) = \{\phi_1, \ldots, \phi_D\}$, where D is the cardinality of the feature subset. This vector forms the input to the classification model that will be described in the next subsection.

D. Linear classifier for hand intent recognition

We train an ensemble of multinomial logistic regression (MLR) models [33] to discriminate between the three categories of motion intent - hand closing, hand opening and relaxing. Each model in the ensemble is trained with a distinct subset of EMG channels. A schematic of the MLR model is shown in Figure 4.

This model consists of three sets of learnable parameters corresponding to the different categories of motion intent - relax, close and open respectively. These parameters are linearly combined with the input feature set $\{\phi_1, \ldots, \phi_D\}$ to give 3 outputs that are passed through a softmax layer to calculate probabilities for each category. Since no non linear activation function except for a softmax is used in the architecture, this model belongs to the class of *linear classifiers*. The output probabilities for each category can be calculated as follows.

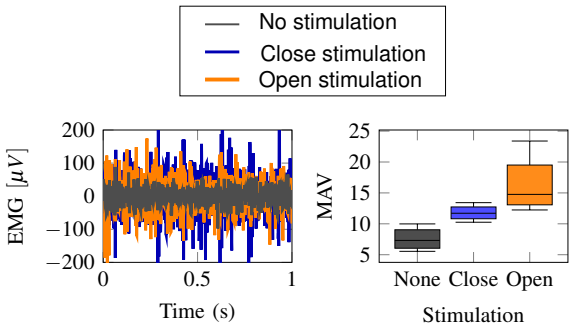


Fig. 5: An illustration of how the filtered sEMG signals vary when the stimulation amplitudes and locations vary. The left column depicts the filtered sEMG signals corresponding to the volitional motion *Open hand* while three different configurations of FES is applied. The right column shows the distribution of the MAV feature for the same configurations via a boxplot. The median, lower and upper quartile of the feature are shown in the box. The whiskers represent the minimum and maximm values of the feature.

$$P^h(\boldsymbol{\phi}) = \frac{\exp[b^h + \sum_{d=1}^{D} w_d^h \cdot \phi_d]}{\sum_{h'=1}^{3} \exp[b^{h'} + \sum_{d=1}^{D} w_d^{h'} \cdot \phi_d]}; h \in \{R, C, O\} \quad (7)$$

where R, C, O respectively denote the categories hand relax, hand close and hand open.

Let the collective set of learnable parameters $\{w_1, \ldots, w_D, b\}^R, \{w_1, \ldots, w_D, b\}^C, \{w_1, \ldots, w_D, b\}^O$ of the MLR model be denoted by the vector $\boldsymbol{\theta}$. These parameters are learned by minimizing the following crossentropy loss over the dataset $\mathscr{D}$ of filtered signal segments with respect to the parameters $\boldsymbol{\theta}$.

$$\mathscr{L}_{\boldsymbol{\theta}}(\mathscr{D}) = -\sum_{\boldsymbol{\phi}, y \in \mathscr{D}} \sum_{h=1}^{3} y^h \log(P^h(\boldsymbol{\phi})); \quad y^h = \begin{cases} 1 & \text{if } y = h \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

E. sEMG acquisition strategy

FES evoked sEMG measurements can differ from sEMG measurements produced by volitional activity even after an adaptive filter has been employed for removing FES-induced artifacts. For each category of volitional hand motion, Figure 5 visually illustrates the differences between sEMG signals that were recorded when no FES was applied and filtered sEMG signals that were recorded while simultaneously applying FES for evoking different hand motions.

As a result of these differences, the distribution of features extracted from the filtered signals where FES was simultaneously applied is different from the case where sEMG is recorded when FES is inactive (also shown in Figure 5). Consequently, a classifier trained only on features from sEMG recordings without FES application may not perform well when used for hand intent recognition for continuous FES control, since in this scenario, the sEMG recordings may be subject to artifacts due to simultaneous application of FES.

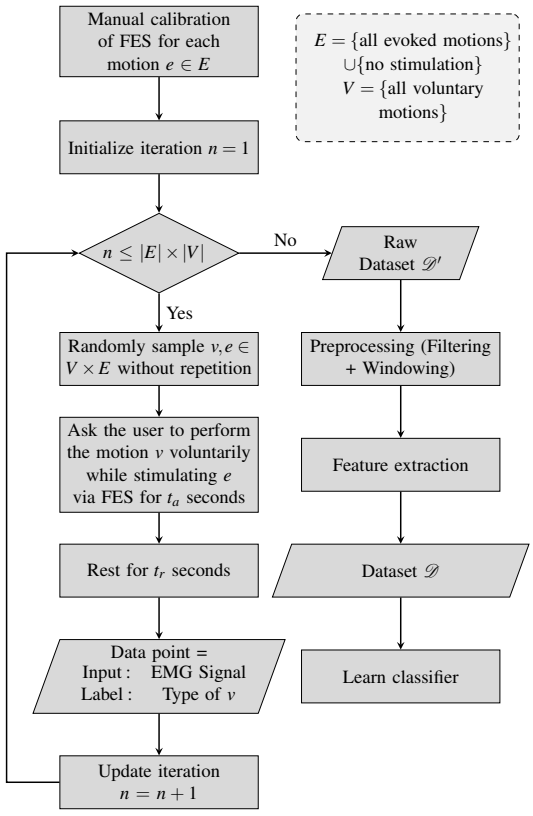


Fig. 6: Framework for learning a hand intent recognition model from limited data for FES-based control. sEMG signals are systematically acquired from a combination of voluntary and evoked motions, processed to yield feature sets and finally used to train a classifier that can identify the volitional intent of the user.

Consequently, we introduce a calibration routine that captures sEMG signals arising from both the user's intentional actions and FES stimulation. The signal combinations to be acquired are crafted such that all possible cases that can be encountered during concurrent FES stimulation are included.

III. Results and Discussion

We perform several analysis to ascertain the validity of the presented approach which we term as MLR-AF (Multinomial Logistic Regression with Adaptive Filter). In addition to comparing it to other approaches, we evaluate the impact of several parameters of the framework. This includes the inclusion of the adaptive filter for preprocessing the data, as well as the impact of including FES evoked sEMG in the training set. Additionally, we examine the impact of the amount of training data on performance, which can affect the user experience of the calibration process, as well as the robustness of the method to partial loss of sEMG information due to possible disconnection of one or more channels.

A. Dataset for evaluation

The sEMG acquisition algorithm described in Figure 6 was implemented in MATLAB and used for collecting the dataset for evaluating the proposed method. The set of volitional motions was given by $V = \{Relax\ hand, Close\ hand, Open\ hand\}$ and that of evoked motions was given by $E = \{No\ Stimulation, Close\ hand, Open\ hand\}$. Prior to the data collection for each participant, FES amplitude for the evoked motions was obtained via manual calibration. The FES amplitude corresponding to *Close hand* varied as 8.8 ± 2.2 mA, while for *Open hand*, the amplitude varied as 9.6 ± 1.8 mA. Each of ten participants were asked to repeat the data collection session 9 times. During each session $|V| \times |E| = 3 \times 3 = 9$ combinations were recorded. Each combination was recorded for 4 seconds. There was a break of 5 seconds in between each session.

B. Evaluation metrics

Several evaluation metrics are used for our analysis. These include macro-averaged precision (MAP), macro-averaged recall (MAR), f1-score (F1) and area under receiver operator characteristic (AUROC). MAP measures fraction of predictions that are correctly assigned to the true class averaged over all classes. MAR describes the fraction of class h examples that are correctly classified, averaged over all classes. F1 is the harmonic mean of MAP and MAR. Finally, AUROC depicts the area under the curve defining ratio of true positive to false positive rate and can be used as a measure of the model's ordering capability, i.e higher values of AUROC indicate that for a randomly chosen example from class h, the output probability of h ranks higher than the other classes. Higher values for all of these metrics indicate better model performance.

Furthermore, we report the significance of classification performance for three different comparisons with a variation of the McNemar's test [34]. The first analysis compares 2 deep learning approaches - convolutional neural network (CNN) and a fully connected neural network (FCN) that directly process the 8 channel sEMG signal to our approach using feature based MLR. The second comparison examines the effect of using the adaptive filter for pre-processing. Finally, we compare a feature-based MLR learned on both volitional and evoked signals to one that is trained only with volitional signals.

For all our evaluations, we used the top 5 hyper-parameters (feature sets for linear classifiers; layer configurations, learning rates and number of training epochs for deep neural networks) for computing the average evaluation metrics and performance statistics. The segment length for calculating the feature sets for MLR was chosen to be $K = 0.2$ seconds. The length of the signal segment for the deep neural network models is also chosen to be $K = 0.2$ seconds. Furthermore, each consecutive pair of sessions was considered to be the training and test session pair. Thus the evaluation metrics were calculated over a concatenation of the disjoint test sessions, and macro-averaged over the participants and hyper-parameters.

TABLE II: Comparative performance of different model architectures and filtering

Model	Adaptive Filter	MAP	MAR	F1	AUROC
FCN-AF	Yes	0.838	0.819	0.820	0.946
CNN-AF	Yes	0.924	0.924	0.923	0.966
MLR-AF	Yes	0.966	0.965	0.965	0.990
FCN	No	0.559	0.525	0.508	0.686
CNN	No	0.585	0.556	0.549	0.702
MLR	No	0.606	0.591	0.581	0.755

C. Comparison with deep learning methods

Deep learning approaches such as convolutional neural networks (CNN), and fully connected neural networks (FCN) are known for extracting features automatically without any feature engineering. However, these networks are also often parameter-dense and therefore need large amounts of data for training the model. We compare these methods to our linear classification approach that uses a fixed cohort of features in Table II. For each model architecture, both cases where an adaptive filter (prefixed by the letters AF) is or is not used for preprocessing were considered.

We note that among the methods using adaptive filtering, our model (MLR) yields the highest values of MAP (0.964), MAR (0.963), F1 (0.963) and AUROC (0.990). We additionally note that when adaptive filtering is used, MLR-AF significantly outperforms FCN-AF ($\chi^2 = 8.125$, p-value< 0.05). However, significance testing of the MLR-AF's performance compared to FCN-AF yielded a p-value> 0.05. Nevertheless, the high performance metrics yielded by MLR-AF indicate that the time-domain features are atleast as good or better than the automatically extracted features produced by the deep learning models. Among methods not using adaptive filtering for data-preprocessing, MLR once again yields the highest values of MAP (0.705), MAR (0.693), F1 (0.689) and AUROC (0.842) and significantly outperforms CNN ($\chi^2 = 5.640$, p-value< 0.05) and FCN ($\chi^2 = 7.845$, p-value$<$ 0.05). One reason for the low performance of the deep-neural networks on unfiltered data can be that the models overfit on the small training dataset and cannot generalize to new data.

We additionally note that comparison of MLR-AF to other shallow networks namely, support vector machines and decision trees yielded no significant differences in performance. For the sake of brevity, a description of these results was excluded in this work.

D. Intent recognition performance with adaptive filtering

We examined the effect of pre-processing the raw sEMG signals with the adaptive filter in our learning framework. The evaluation results are shown in Table II. When adaptive filtering was included during data pre-processing, the performance of the linear classification model (MLR-AF) was noted to be significantly better ($\chi^2 = 8.68$, p-value< 0.05) than when it was not (MLR). This is also true for both

TABLE III: Impact of including/excluding FES evoked sEMG during training.

Training Dataset	MAP	MAR	F1	AUROC
Volitional+FES-evoked sEMG	0.966	0.965	0.965	0.990
Volitional sEMG	0.943	0.942	0.943	0.986

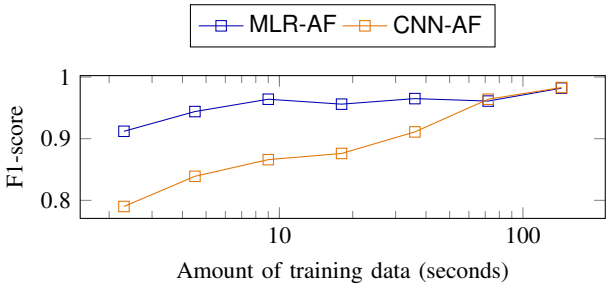


Fig. 7: Comparative performance of MLR-AF with CNN-AF for varying amounts of training data.

the deep learning approaches, where CNN-AF demonstrated significant performance improvement over CNN ($\chi^2 = 8.48$, p-value< 0.05). The same held true for the fully connected architecture when FCN-AF outperformed FCN ($\chi^2 = 8.01$, p-value< 0.05). This indicates that the raw sEMG signal is affected by artifacts that dominate the feature extraction and render the extracted features less informative for hand intent recognition, both for engineered features and deep neural networks like CNNs and FCNs. Filtering these effects to some extent before training the classifier is necessary for robust hand intent recognition.

E. Inclusion of FES evoked signals in the training set

We examine the impact of including combinations of volitional and evoked sEMG responses instead of purely volitional sEMG without external stimulation. Table III lists the average performance metrics for both conditions. We note that the linear classification model trained with both volitional and evoked responses yields the highest values of metrics MAP (0.966), MAR (0.965), F1 (0.965) and AUROC (0.990). Significance testing of this model compared to the model trained only on volitional sEMG data yields a p-value$= 0.05$ ($\chi^2 = 3.735$). This indicates that the adaptive filter may have been unable to remove all FES-induced artifacts from the sEMG signals, affecting model performance adversely. For continuous FES-based assistive control, decreased model performance could impact user safety due to inaccurately estimated hand intentions, leading to unwanted stimulation. Including sEMG samples affected by FES stimulation in the training dataset can mitigate this.

F. Training data sufficiency

The amount of data necessary to train a model for hand intent recognition regulates the speed of the calibration routine to some extent and in turn impacts the user's experience. We therefore attempt to infer the amount of training data

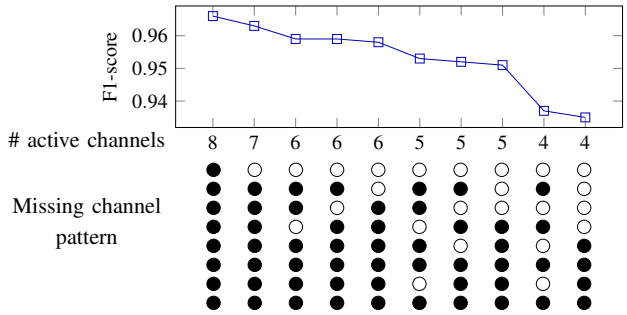


Fig. 8: MLR-AF robustness for different patterns of missing sEMG channels. The empty circles indicate the inactive/missing sEMG channels.

that is sufficient for learning a robust hand intent recognition model. Since each session itself consists of 9 sets of 4 second long recordings, we simulate longer periods of training data by concatenating data from consecutive sessions and using the last (ninth) session for performance evaluation. Figure 7 compares the F1-score for varying amounts of training data for MLR-AF and the deep-network model CNN-AF. MLR-AF's performance was noted to increase drastically with the amount of training data for recording times≤ 7.2, after which its value remained steady. High F1-scores (> 0.95) are observed for MLR-AF for recording times≥ 7.2 seconds, indicating that an accurate classifier can be learned with a small dataset. In contrast, equivalent F1-scores are achieved by the highest performing deep network model CNN-AF only with higher amounts of data (≥ 72 seconds).

G. Robustness to missing channels

Finally, we compare the robustness of our approach to loss of information from one more sEMG channels. This can occur due to wear and tear of the sEMG array or lack of sufficient adhesive on the electrode surface, which can result in the electrode to lose contact with the user's skin. We report the F1-score for different channel configuration in Figure 8. The availability of data from more channels imply the higher probability of detecting different muscle responses. With all 8 channels available, the highest values for the F1-score (0.966) is achieved. This performance tends to steadily decrease with the loss of more channels. Furthermore, we note that when lost channels are clustered together, i.e data from a large region is missing, the classification performance is worse. A drastic decrease in performance is noted when four channels of data are missing, since this implies the potential loss of signals from half of the covered region.

H. Future work

The proposed method for hand-intent recognition has been successfully validated on healthy participants, demonstrating its potential for automatically recognizing hand intent in stroke patients. However, analysis of gesture-specific sEMG data from stroke patients presents additional challenges that arise from the reduced number of active motor units and the altered myoelectric patterns typical in stroke patients [35]. To

better understand these issues and evaluate their impact on the effectiveness of our approach, we are currently collecting additional data for a future study.

IV. Conclusion

In this work, we proposed a framework for learning a hand intent recognition model that can be used for FES-based assistive control from limited data. sEMG recordings resulting from volitional intent and FES are systematically collected for each possible combination of volitional intent and FES input. This data is used for training a linear classifier with time-domain features to robustly identify hand motion primitives in the presence of FES. Our feature based model outperformed deep neural networks like CNN and FCN, a possible result of the latter models overfitting on the training data. We further show that our model is able to learn from limited amounts of data, therefore allowing for quick training of this model from scratch for individual users.

References

[1] L. Mercier, T. Audet, R. Hébert, A. Rochette, and M.-F. Dubois, "Impact of motor, cognitive, and perceptual disorders on ability to perform activities of daily living after stroke," *Stroke*, vol. 32, no. 11, pp. 2602–2608, 2001.

[2] A. García Álvarez, A. Roby-Brami, J. Robertson, and N. Roche, "Functional classification of grasp strategies used by hemiplegic patients," *PLoS One*, vol. 12, no. 11, p. e0187608, 2017.

[3] F. Quandt and F. C. Hummel, "The influence of functional electrical stimulation on hand motor recovery in stroke patients: a review," *Experimental & translational stroke medicine*, vol. 6, no. 1, pp. 1–7, 2014.

[4] H. PETER, "Functional electrical stimulation in neurorehabilitation," *Textbook of Neural Repair and Rehabilitation*, pp. 119–135, 2006.

[5] L. Y. Chang, N. S. Pollard, T. M. Mitchell, and E. P. Xing, "Feature selection for grasp recognition from optical markers," in *2007 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2007, pp. 2944–2950.

[6] Y. Herbst, L. Zelnik-Manor, and A. Wolf, "Analysis of subject specific grasping patterns," *Plos one*, vol. 15, no. 7, p. e0234969, 2020.

[7] M. Štrbac, S. Kočović, M. Marković, D. B. Popović *et al.*, "Microsoft kinect-based artificial perception system for control of functional electrical stimulation assisted grasping," *BioMed research international*, vol. 2014, 2014.

[8] L. T. Taverne, M. Cognolato, T. Bützer, R. Gassert, and O. Hilliges, "Video-based prediction of hand-grasp preshaping with application to prosthesis control," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 4975–4982.

[9] J. de Vries, A. van Ommeren, G. Prange-Lasonder, J. Rietman, and P. Veltink, "Detection of the intention to grasp during reach movements," *Journal of Rehabilitation and Assistive Technologies Engineering*, vol. 5, p. 2055668317752850, 2018.

[10] W. Yu, H. Yamaguchi, H. Yokoi, M. Maruishi, Y. Mano, and Y. Kakazu, "EMG automatic switch for fes control for hemiplegics using artificial neural network," *Robotics and Autonomous Systems*, vol. 40, no. 2-3, pp. 213–224, 2002.

[11] C. Salchow, A. Dorn, M. Valtin, and T. Schauer, "Intention recognition for fes in a grasp-and-release task using volitional EMG and inertial sensors," *Current directions in biomedical engineering*, vol. 3, no. 2, pp. 161–165, 2017.

[12] C. L. Toledo-Peral, J. Gutiérrez-Martínez, J. A. Mercado-Gutiérrez, A. I. Martín-Vignon-Whaley, A. Vera-Hernández, L. Leija-Salas *et al.*, "sEMG signal acquisition strategy towards hand fes control," *Journal of Healthcare Engineering*, vol. 2018, 2018.

[13] I. Batzianoulis, N. E. Krausz, A. M. Simon, L. Hargrove, and A. Billard, "Decoding the grasping intention from electromyography during reaching motions," *Journal of neuroengineering and rehabilitation*, vol. 15, no. 1, pp. 1–13, 2018.

[14] E. Trigili, L. Grazi, S. Crea, A. Accogli, J. Carpaneto, S. Micera, N. Vitiello, and A. Panarese, "Detection of movement onset using EMG signals for upper-limb exoskeletons in reaching tasks," *Journal of neuroengineering and rehabilitation*, vol. 16, pp. 1–16, 2019.

[15] D. C. Irimia, M. S. Poboroniuc, F. Serea, A. Baciu, and R. Olaru, "Controlling a fes-exoskeleton rehabilitation system by means of brain-computer interface," in *2016 International Conference and Exposition on Electrical and Power Engineering (EPE)*. IEEE, 2016, pp. 352–355.

[16] B. C. Osuagwu, L. Wallace, M. Fraser, and A. Vuckovic, "Rehabilitation of hand in subacute tetraplegic patients based on brain computer interface and functional electrical stimulation: a randomised pilot study," *Journal of neural engineering*, vol. 13, no. 6, p. 065002, 2016.

[17] M.-S. Loh, Y.-C. Kuan, C.-W. Wu, C.-D. Liao, J.-P. Hong, and H.-C. Chen, "Upper extremity contralaterally controlled functional electrical stimulation versus neuromuscular electrical stimulation in post-stroke individuals: A meta-analysis of randomized controlled trials," *Neurorehabilitation and Neural Repair*, vol. 36, no. 7, pp. 472–482, 2022.

[18] S. Balasubramanian, E. Garcia-Cossio, N. Birbaumer, E. Burdet, and A. Ramos-Murguialday, "Is EMG a viable alternative to bci for detecting movement intention in severe stroke?" *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 12, pp. 2790–2797, 2018.

[19] A. Phinyomark, C. Limsakul, and P. Phukpattaranont, "A novel feature extraction for robust EMG pattern recognition," *arXiv preprint arXiv:0912.3973*, 2009.

[20] B. Hudgins, P. Parker, and R. N. Scott, "A new strategy for multifunction myoelectric control," *IEEE transactions on biomedical engineering*, vol. 40, no. 1, pp. 82–94, 1993.

[21] D. Tkach, H. Huang, and T. A. Kuiken, "Study of stability of time-domain features for electromyographic pattern recognition," *Journal of neuroengineering and rehabilitation*, vol. 7, no. 1, pp. 1–13, 2010.

[22] Y. Narayan, "SEMG signal classification using knn classifier with fd and tfd features," *Materials Today: Proceedings*, vol. 37, pp. 3219–3225, 2021.

[23] C. Sapsanis, G. Georgoulas, and A. Tzes, "EMG based classification of basic hand movements based on time-frequency features," in *21st Mediterranean conference on control and automation*. IEEE, 2013, pp. 716–722.

[24] P. Tsinganos, B. Cornelis, J. Cornelis, B. Jansen, and A. Skodras, "Deep learning in EMG-based gesture recognition." in *PhyCS*, 2018, pp. 107–114.

[25] S. Tam, M. Boukadoum, A. Campeau-Lecours, and B. Gosselin, "A fully embedded adaptive real-time hand gesture classifier leveraging hd-sEMG and deep learning," *IEEE transactions on biomedical circuits and systems*, vol. 14, no. 2, pp. 232–243, 2019.

[26] S. Sennels, F. Biering-Sorensen, O. T. Andersen, and S. D. Hansen, "Functional neuromuscular stimulation controlled by surface electromyographic signals produced by volitional activation of the same muscle: adaptive removal of the muscle response from the recorded EMG-signal," *IEEE Transactions on Rehabilitation Engineering*, vol. 5, no. 2, pp. 195–206, 1997.

[27] D. Zhang and W. T. Ang, "Reciprocal EMG controlled fes for pathological tremor suppression of forearm," in *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2007, pp. 4810–4813.

[28] H. Yeom, Y. Park, Y. Yoon, T. Shin, and H. Yoon, "An adaptive m-wave canceler for the EMG controlled functional electrical stimulator and its fpga implementation," in *The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 2, 2004, pp. 4122–4125.

[29] R. Pilkar, M. Yarossi, A. Ramanujam, V. Rajagopalan, M. B. Bayram, M. Mitchell, S. Canton, and G. Forrest, "Application of empirical mode decomposition combined with notch filtering for interpretation of surface electromyograms during functional electrical stimulation," *IEEE transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 8, pp. 1268–1277, 2016.

[30] X. Chen, Y. Jiao, D. Zhang, Y. Wang, X. Wang, Y. Zang, Z. Liang, and P. Xie, "An adaptive spatial filtering method for multi-channel EMG artifact removal during functional electrical stimulation with time-variant parameters," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2023.

[31] B. A. Osuagwu, E. Whicher, and R. Shirley, "Active proportional

electromyogram controlled functional electrical stimulation system," *Scientific reports*, vol. 10, no. 1, p. 21242, 2020.

[32] H. Kavianirad, S. Endo, T. Keller, and S. Hirche, "EMG-based volitional torque estimation in functional electrical stimulation control," in *2022 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES)*. IEEE, 2022, pp. 171–176.

[33] C. M. Bishop and N. M. Nasrabadi, *Pattern recognition and machine learning*. Springer, 2006, vol. 4, no. 4.

[34] V. L. Durkalski, Y. Y. Palesch, S. R. Lipsitz, and P. F. Rust, "Analysis of clustered matched-pair data," *Statistics in medicine*, vol. 22, no. 15, pp. 2417–2428, 2003.

[35] X. Li, J. Liu, S. Li, Y.-C. Wang, and P. Zhou, "Examination of hand muscle activation and motor unit indices derived from surface emg in chronic stroke," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 12, pp. 2891–2898, 2014.