Edinburgh Research Explorer

Decoding HD-EMG Signals for Myoelectric Control-How Small Can the Analysis Window Size be?

Spatio-Temporal Warping for Myoelectric Control: An Offline, Feasibility Study

Milad Jabbari, Rami Khushaba, Kianoush Nazarpour

Abstract.

Objective: The efficacy of an adopted feature extraction method directly affects the classification of the electromyographic (EMG) signals in myoelectric control applications. Most methods attempt to extract the dynamics of the multi-channel EMG signals in the time domain and on a channel-by-channel, or at best pairs of channels, basis. However, considering multi-channel information to build a similarity matrix has not been taken into account.

Approach: Combining methods of long and short-term memory (LSTM) and dynamic temporal warping (DTW), we developed a new feature, called spatio-temporal warping (STW), for myoelectric signals. This method captures the spatio-temporal relationships of multi-channels EMG signals.

Main Results: Across four online databases, we show that in terms of average classification error and standard deviation values, the STW feature outperforms traditional features by 5% to 17%. In comparison to the more recent deep learning models, e.g. convolutional neural networks (CNN), STW outperformed by 5% to 18%. Also, STW showed enhanced performance when compared to the CNN+LSTM model by 2% to 14%. All differences were statistically significant with a large effect size.

Significance: This feasibility study provides evidence supporting the hypothesis that spatio-temporal warping of the EMG signals can enhance the classification accuracy in an explainable way when compared to recent deep learning methods. Future work includes real-time implementation of the method and testing for prosthesis control.

1. Introduction

The electromyographic (EMG) signal, recorded from the stump muscles, is a valuable, rich, complex, and dynamic source of information for prosthesis control. Research shows that providing natural EMG-based feedback with different approaches including grasp/wrist force and movement estimation [1–4], continuous finger trajectory decoding [5], and discrete movement classification [6, 7] could achieve promising results for desirable control of the prosthesis. Machine learning is a candidate tool in mapping motor intent to prosthesis control [8–10]. Extracting reliable features from the EMG signals plays a vital role in the control of upper limb prostheses with pattern recognition [8,9,11]. A wide range of factors affect the accuracy of machine learning-based methods [12–17]. The separability of the features can have a greater influence on the performance of EMG decoding than the type of the classifier [12–14]. Therefore, research on the development and real-time testing of more robust features is on-going.

Current EMG feature extraction methods suffer from two intrinsic limitations:

 (i) they are cross-sectional [18–21]; that is, they cannot extract the inter-temporal dependencies that exist between feature extraction windows;

(ii) they merely concatenate features extracted from individual channels [15, 22–24]; not capturing the synergistic and spatial patterns of muscles activity;

These limitations are depicted in the left side of figure 1. Therefore, the development of feature extraction methods that can capture the temporal dynamics of the EMG signals in a spatially-aware way, such as the one shown on the right side of figure 1, has received increased attention recently [15, 21, 25–38].

To address the former, many hand-crafted features [15, 25–27] and deep learning-based approaches [21, 28–30] are introduced. The fusion of time-domain descriptors (fTDD) and deep Long Short-Term Memory (LSTM) networks yield higher classification accuracy than the conventional time-domain (TD) features [19, 21, 39, 40]. These methods excel in extracting the temporal information by concatenating features. Hence, they are blind to spatial information.

To address the second limitation, deep learning methods, e.g. convolutional neural networks (CNN), construct spatial feature sets and capture the relationship between EMG channels [31–35]. Convolutional structures do not typically consider within-channel inter-temporal information of the EMG signals. Therefore, there has been a surge of research in combining CNNs and LSTM to capture the spatio-temporal features of the EMG signals [36–38]. However, the prohibitively large number of model parameters, e.g. 34-95k in [41], 104k in [32], and 30-549k in [42], may impose a barrier in real-time implementation for myoelectric control. Traditional features have been also used to extract spatial information [15, 23, 43, 44] but that conversely leads to the loss of temporal information.

Hand-crafted features such as the temporal-spatial descriptors (TSD) [15], fTDD [19], and time domain with auto-regressive model parameters (TDAR) have been
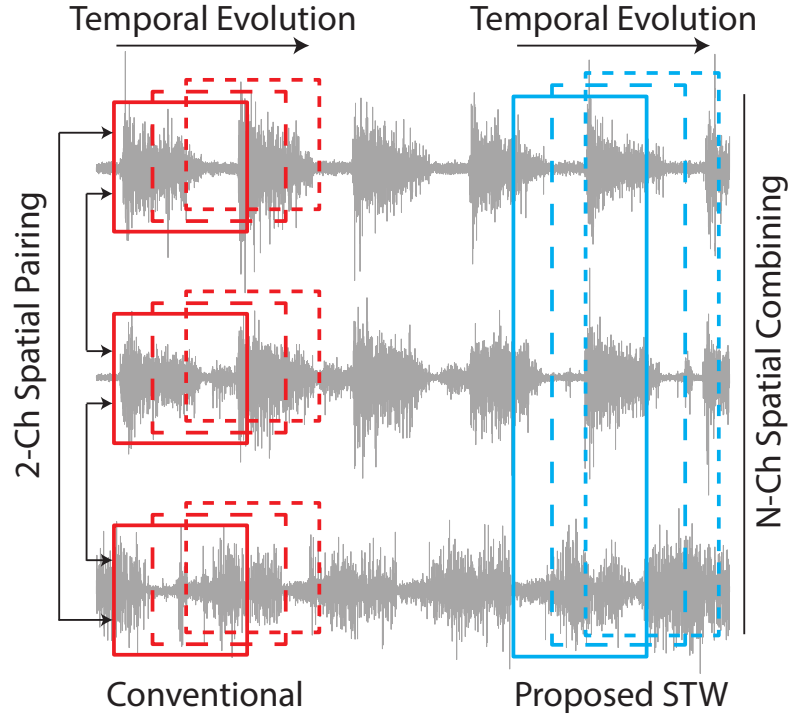
Figure 1. The conventional EMG feature extraction approach that suffers from the lack of inter-temporal focus, between windows, and the lack of spatial focus, between channels (left); The proposed STW approach with across-channel feature extraction and considering the inter-temporal relationships between windows (right).

recently used to complement different CNN and LSTM models [21, 45, 46]. Hence, they can facilitate the adoption of recent deep learning models for real-time clinical implementations given their reduced computational burden in comparison to deep learning models.

We address both limitations with a novel feature extraction method that outperforms deep learning methods. Dynamic time warping was employed to efficiently capture the nonlinear similarity between the EMG signals. For the temporal aspect, we developed a novel feature, named Spatio-Temporal Warping (STW). Figure 1 summarises the challenge and the proposed approach.

2. Methods

Our STW approach has two building blocks. The first is the Dynamic Time Warping (DTW) algorithm [47]. The second component temporally fuses the extracted DTW features with that of a previous time window and that of a global trend. Figure 2 depicts this method.

Figure 2. Flow diagram for the proposed STW extraction algorithm.

2.1. Dynamic Time Warping (DTW)

The DTW method measures the similarity between two sequences of different temporal dynamics. It has been used in speech recognition [47], data mining [48], gesture recognition [49], robotics [50] and myoelectric control [51–53]. Previous research utilizing DTW in myoelectric control considered time-series similarity across training and testing EMG sequences from individual channels only. Using DTW, we created a similarity matrix between all pairs of EMG signals. DTW can also be applied across a set of signals together.

Assume two temporal sequences of length l: $\mathbf{a} = \{a_1, a_2, \cdots, a_l\}$ and $\mathbf{b} = \{b_1, b_2, \cdots, b_l\}$, representing two EMG signals. We define $\mathbf{D}(\mathbf{a}, \mathbf{b})$ as an $l \times l$ distance matrix between $\mathbf{a}$ and $\mathbf{b}$, with $\mathbf{D}_{ij} = (a_i - b_j)^2$. In the case of Euclidean distance, we generally set $i = j$, that is the distance along the same index from the two time series. The path for the Euclidean distance is hence the one along the diagonal of the matrix $\mathbf{D}$, where $i = j$. Although, due to its simplicity and efficiency, Euclidean distance is widely used in different approaches, it suffers from two main problems: first, it is not practical for situation that the length of the sequences is not equal and second, it is sensitive to time shifting. Finding an optimal path using DTW can address these challenges. In the case of DTW method, the distance is calculated through a

different warping path $\mathcal{P}$ generated by traversing the matrix $\mathbf{D}$ along ordered pairs of positions: $\mathcal{P} =< (e_1, f_1), (e_2, f_2), \cdots, (e_l, f_l) >$ where $e_i \in [1 : l]$ and $f_i \in [1 : l]$ are the positions that make the warping path. A valid warping path must satisfy the conditions $(e_1, f_1) = (1, 1)$, $(e_l, f_l) = (l, l)$, $0 \leq e_{i+1} - e_i \leq 1$ and $0 \leq f_{i+1} - f_i \leq 1$ for all $i < l$.

A warping limit is typically imposed on the distances, that is $|e_i - f_i| \leq w.l$, $\forall (e_i, f_i) \in P^*$. The value of w is the maximum allowed warping path to deviate from the diagonal. According to [54], the distance D to any path $\mathcal{P}$ is given with $D_{\mathcal{P}}(\mathbf{a}, \mathbf{b}) = \sum_{i=1}^{s} p_i$, where $p_i = \mathbf{D}_{e_i, f_i}$ is the distance between element at position e_i of $\mathbf{a}$ and at position f_i of $\mathbf{b}$ for the ith pair of points in a proposed warping path P. Considering a space of all possible paths as $\mathcal{P}$, then the DTW path P^* is the one that has the minimum distance: $P^* = \min_{P \in \mathcal{P}} \mathbf{D}_{\mathcal{P}}(\mathbf{a}, \mathbf{b})$.

We utilized the aforementioned distances between all NC EMG signals, i.e. $(NC \times (NC - 1))/2$ pairs, as shown in figure 2. Inspired by [19, 25] the DTW features of all pairs of first and second derivatives of EMG signals were also computed. Hence, our method extracts $3 \times (NC \times (NC - 1))/2$ features, in total. All calculated DTW distances were then concatenated to form a features vector and passed through logarithmic scaling for improved accuracy, similar to [55].

2.2. Spatio-Temporal Warping (STW)

We first multiplied the features vector extracted from the current windows, DTW_t, with that extracted from the previous window DTW_{t-n}. To account for the long-term memory, our algorithm borrowed the cell state (c) concept of the LSTM method. This was the information highway that passes through all the LSTM cells across all time steps to evolve a recursive representation of the features. It was therefore continuously updated within each cell. At the same time, the contents of each cell were updated with a weighted contribution of the cell state through the parameter β. The selection of the parameter β was empirically made with the objective of achieving the best classification performance across several datasets.

$$f_t = DTW_t \odot DTW_{t-1} + \beta c \tag{1}$$

where f_t representing the extracted features from the current time t, $\odot$ denotes the element-wise multiplication operation, and c symbolises the cell state.

The STW algorithm also presented a dynamic normalization mechanism; borrowing it from neural machine translation [56]. Attention mechanism provides modeling of dependencies without considering their distance in the input or output sequences. We used this method within each cell to normalize the current windows features and before adding it to the aforementioned features to normalize the long-term memory component. The final features representation is

$$f_{ti} = \frac{f_{ti}}{\sum_i f_{ti}} + \log \left(1 + \frac{c_i}{\sum_i c_i} \right) \tag{2}$$

where subscript i indicates the normalization applied across features at time t.

Table 1. Details of the utilized EMG datasets.

Dataset	Subjects	Channels	Samp.Freq (Hz)	Classes
DB1	27	10 (Otto Bock)	100	52
DB5	10	16 (2 MYOs)	200	53
DB7	22	12	2000	41
3DC	22	10 (3DC), 8 (MYO)	1000 (3DC), 200 (MYO)	11
SC-EMG	5	10	1000	7

2.3. EMG Data sets

Five EMG datasets were utilized to test the performance of the proposed feature extraction algorithm as described in table 1.

We used database-1 (DB1), database-5 (DB5), and database-7 (DB7) of the Ninapro repository [57–60], comprising respectively, 27 limbed-intact subjects (52 movements), 10 limb-intact subjects (52 movements and rest) and 20 limb-intact and two amputees (40 movements and rest). EMG signals were recorded using 10 Otto Bock electrodes and two wearable MYO armbands on the forearm in DB1 and DB5, respectively. In DB7, 12 Trigno (Delsys, USA) EMG sensors were placed on the limb. The third EMG dataset [41] included data from 22 limb-intact subjects each performing eleven hand/wrist gestures using two armbands. The fourth dataset, that is selective classification (SC-EMG), included data from 5 traumatic long trans-radial amputees [61], each performing seven classes of motion. Ten electrodes were placed around the forearm at the point of the largest diameter.

2.4. Feature extraction

The below features were extracted from windows of 150 ms at 50 ms increments:

- **HTD**: Hudgins' TD feature set [18], comprising: mean absolute value (MAV), MAV slope (MAVS), zero crossings (ZC), slope sign changes (SSC), and waveform length (WL);

- **AR-RMS**: The 6th-order auto-regressive (AR) coefficients and the RMS;

- **LSF9**: The lower sampling rate features defined in [22], comprising L-scale, maximum fractal length, the mean value of the square root, Willison amplitude, ZC, RMS, integrated absolute value, difference absolute standard deviation value, and variance;

- **ATD**: Combined AR with TD as defined in [60], comprising MAV, WL, 4th-order AR coefficients and log-variance (LogVar),

- **STFS**: Spatio-temporal features from [62], comprising integral square descriptor, normalized root-square coefficient of first and second differential derivatives, mean log-kernel, an estimate of mean derivative of the higher-order moments per sliding window, and a measure of spatial muscle information.

- **fTDD**: the fusion of TD features from [19] with six features representing the first 3 even power spectrum moments, with an irregularity factor, a sparsity measure, and the ratio of WL of the first derivative to that of the second derivative. The step size was 15.

- **TSD**: The temporal spatial-descriptors from [15].

To ensure a fair comparison, the dimensionality of all feature sets was reduced to $c - 1$, where c is the number of classes, using the spectral regression feature projection method [63].

2.5. Classifiers

The following classifiers were chosen: linear discriminant analysis (LDA), extreme learning machine (ELM), k-nearest neighbor (KNN), and support vector machines (SVM). In the ELM classifier, one hidden layer with 1250 neurons was utilized. The parameters of SVM were optimized for each dataset, as the performance of SVM is susceptible to the kernel function parameter γ and the regularization parameter C, while for KNN, the parameter K was empirically set to $K = 5$. Additionally, three deep learning models were also implemented; LSTM, CNN, and a combination of CNN and LSTM. These models are shown in figure 3, designed empirically to reflect the best classification accuracy results using the raw EMG data. Raw EMG inputs to models like CNN (or even CNN+LSTM) are more successful than the spectrogram images [31, 41]. The root-mean-square (RMS) of raw EMG signals were therefore calculated, generating NC scalar values for each analysis window of 150 ms. These values were then turned into pseudo-images by multiplying each generated vector of size ($NC \times 1$) by its transpose, resulting in an $NC \times NC$ images. The images were scaled logarithmically and then provided as inputs to the CNN and CNN+LSTM models. For the LSTM model, the raw EMG samples for each 150 ms window of the NC channels were used as features.

2.6. Statistical Analysis

The Wilcoxon signed rank test was applied to verify the statistical significance of the achieved results, with the results being considered significant for a p-value smaller than 0.05. Finally, the size of differences observed was measured using Cohen's effect size d for paired samples defined as the difference between two group means divided by the standard deviation [22]. A set of predefined thresholds of 0.2, 0.5, 0.8, 1.2, and >2 are usually employed to equate the effect size to small, medium, large, very large, and huge effects respectively. Additionally, when reporting Cohen's d value for testing one method versus another, the direction of the effect was calculated by subtracting the mean of the latter from that of the former. MATLAB 2020a was utilized for all experiments on a laptop with an i7 processor, 16 GB of RAM, and a GPU unit (NVIDIA GeForce RTX 2060).

3. Results

3.1. Results of DB5 Dataset

The average classification error rates across all folds and subjects are shown in figure 4(A). Furthermore, the corresponding bar plots with standard deviation values are shown in figure 1 of Supplementary Materials. In terms of the deep models, CNN+LSTM (noted with CNNL, for brevity) showed a significantly better performance than the individual models of CNN and LSTM ($p < 0.001$ for both tests, $d = 3.15$ for CNN vs. CNN+LSTM and $d = 3.55$ for LSTM vs. CNN+LSTM). CNN also significantly outperformed LSTM ($p < 0.001, d = 2.21$). This finding indicates that the temporal-spatial information captured by CNN+LSTM extracts more information than the individual CNN and LSTM, while the spatial information captured by CNN appears more important on this dataset than the temporal information captured by LSTM.

In terms of the traditional hand-crafted feature extraction algorithms, as it is clear in figure 4(A), HTD and ARRMS were the worst performers on the 53 class problem in DB5, with all other methods significantly outperforming them, except for LSTM. When comparing HTD and LSTM, no difference was found ($p = 0.517, d = 0.10$), which is in line with the findings in [21].

We observed significant differences between LSF9 and ATD ($p = 0.030, d = 0.26$), and large differences between LSF9 and STFS ($p < 0.001, d = 0.80$) and LSF9 and TSD ($p < 0.001, d = 0.80$). These results also show that fTDD achieved significantly lower error rates than all other traditional and deep learning feature extraction methods, including CNN+LSTM, except the proposed STW features that significantly outperformed fTDD ($p < 0.001, d = 2.64$). In contrast, STW significantly outperformed all other feature extraction and/or learning methods, including CNN+LSTM, with all tests having $d > 2$. Overall, the proposed approach yielded an average decrease in classification error of 19.81%±4.36% in comparison to all other methods considered in this work. Additionally, it showed a significantly lower standard deviation.

In addition to the classification accuracy metric, we evaluated the proposed STW method in comparison with other hand-crafted feature extraction algorithms using three other measurements including Fscore, Recall, and Precision [64]. Results are illustrated in figure 3 of Supplementary Materials. Average values for Fscore, Recall, and Precision, as well as statistical tests, prove that STW can outperform all conventional hand-crafted algorithms. Furthermore, to perform a quantitative comparison; in terms of separability of clusters, between STW and other conventional feature extraction methods, Davies-Bouldin Index (DBI) was used [65]. Averaged DBI values calculated across test folds and all subjects are shown in figure 4 of Supplementary Materials. As it is visible clearly, STW has achieved the lowest DBI values which show its highest capability in feature discrimination criteria in comparison with other feature extraction methods.
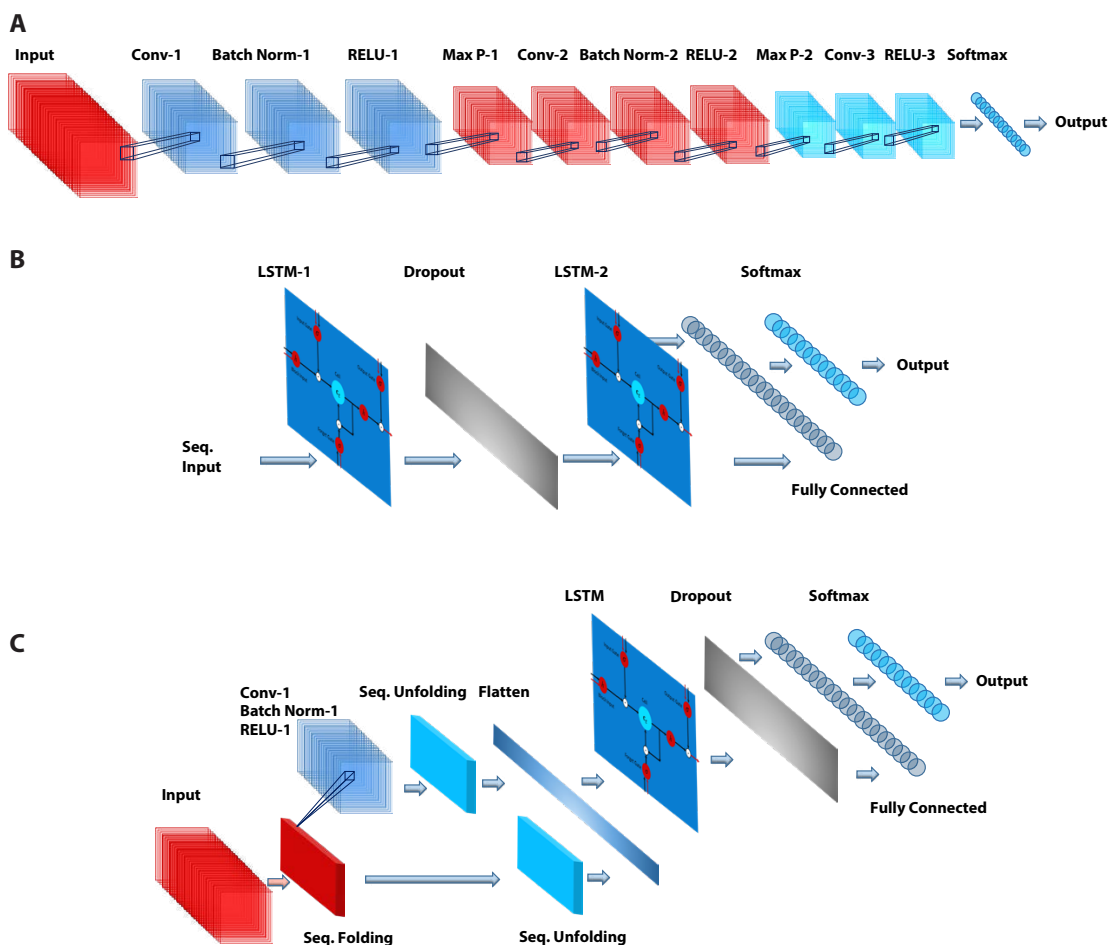
Figure 3. The schematic architectures of the utilized deep learning models: A. CNN model including two convolutional layers; B. LSTM network consists of two LSTM layers; C. CNN+LSTM structure with one convolutional layer and one LSTM layer. The output of all models is connected to a softmax layer.

3.2. Results of the SC-EMG Datasets

The average classification error results on the SC-EMG datasets are shown in figure 4(B) for all traditional hand-crafted and deep learning models (Bar plots of the results are provided in figure 1 of Supplementary Materials). In terms of the deep models, it is interesting to see that, LSTM performed better than CNN in terms of the average error rates across all amputees. This could be in part attributed to the fact that the amputation affects muscles morphology and synergy and hence the performance of CNN is impacted by its capabilities in capturing spatial information between signals which may vary (i.e., warp) in timing as a result of amputation. In this direction, a recent work by Tallec and Ollivier [66] shows that LSTM networks have the capability to learn to warp input sequences. Despite the differences in the average classification errors of deep models, the Wilcoxon signed rank test though revealed no statistically significant difference between LSTM and CNN, LSTM and CNN+LSTM, and neither

between CNN and CNN+LSTM ($p > 0.05$ for all tests). In terms of the hand-crafted methods, fTDD performed significantly worse than TSD ($p < 0.001$) while the proposed STW significantly outperformed all other hand-crafted and deep models ($d = 2.54, 2.09, 1.87, 2.08, 1.87, 0.89$, and 2.05 for STW vs. HTD, AR-RMS, STFS, LSF9, ATD, TSD, and fTDD, respectively, and $p < 0.001$ for all comparisons). The performance of the STW method can be further justified by its ability to approximate the warping path that best aligns the two signals of interest in the time domain, supported by the DTW stage. Such a path is a more representative of the distance between the signals than what Euclidean distance can offer; despite nonlinear variations in speed. The interested reader in referred to [47] where the details of the DTW method can be found.

3.3. Results of DB7 Datasets

The same pattern of results was observed for DB7 datasets. Average classification error rates are shown in figure 4(C) as well as corresponding bar plots in figure 1 of Supplementary Materials. To avoid repetition, we used only an LDA classifier for decoding. HTD and AR-RMS were performed worst. In comparison to HTD and AR-RMS, LSTM performed better than both methods on DB7 datasets ($p < 0.01$ for both tests, $d = 2.06$ for HTD vs. LSTM, $d = 1.15$ for AR-RMS vs. LSTM). On the other hand, both CNN+LSTM and CNN significantly outperformed LSTM ($p < 0.001$ for both tests, $d = 2.97$ for LSTM vs. CNN+LSTM and $d = 2.36$ for LSTM vs. CNN), with CNN+LSTM also significantly outperforming CNN ($p < 0.001, d = 1.18$). On the other hand, both fTDD and the proposed STW significantly outperformed all other methods, including CNN+LSTM, while STW also significantly outperforming fTDD ($p < 0.001, d = 1.32$).

3.4. Results of the 3DC Dataset

The average classification values achieved for the 3DC dataset are illustrated in figure 4(D), and (E) as well as the corresponding bar plots in figure 1 of Supplementary Materials. As two different armbands including MYO and 3DC were used in this dataset, separate analyses were conducted. Instead of raw EMG signals, extracted STW features were fed into the deep models. KNN achieved the lowest average classification error, but statistical tests reveal no significant differences. Similar to the SC-EMG dataset, LSTM could achieve the best performance compared to CNN+LSTM, and CNN in terms of average classification error rates. Wilcoxon signed rank test revealed significant differences between LSTM and CNN and also between LSTM and CNN+LSTM ($p < 0.001$). With respect to the 3DC armband, similar results were obtained as LSTM significantly outperformed CNN and CNN+LSTM ($p < 0.01$).
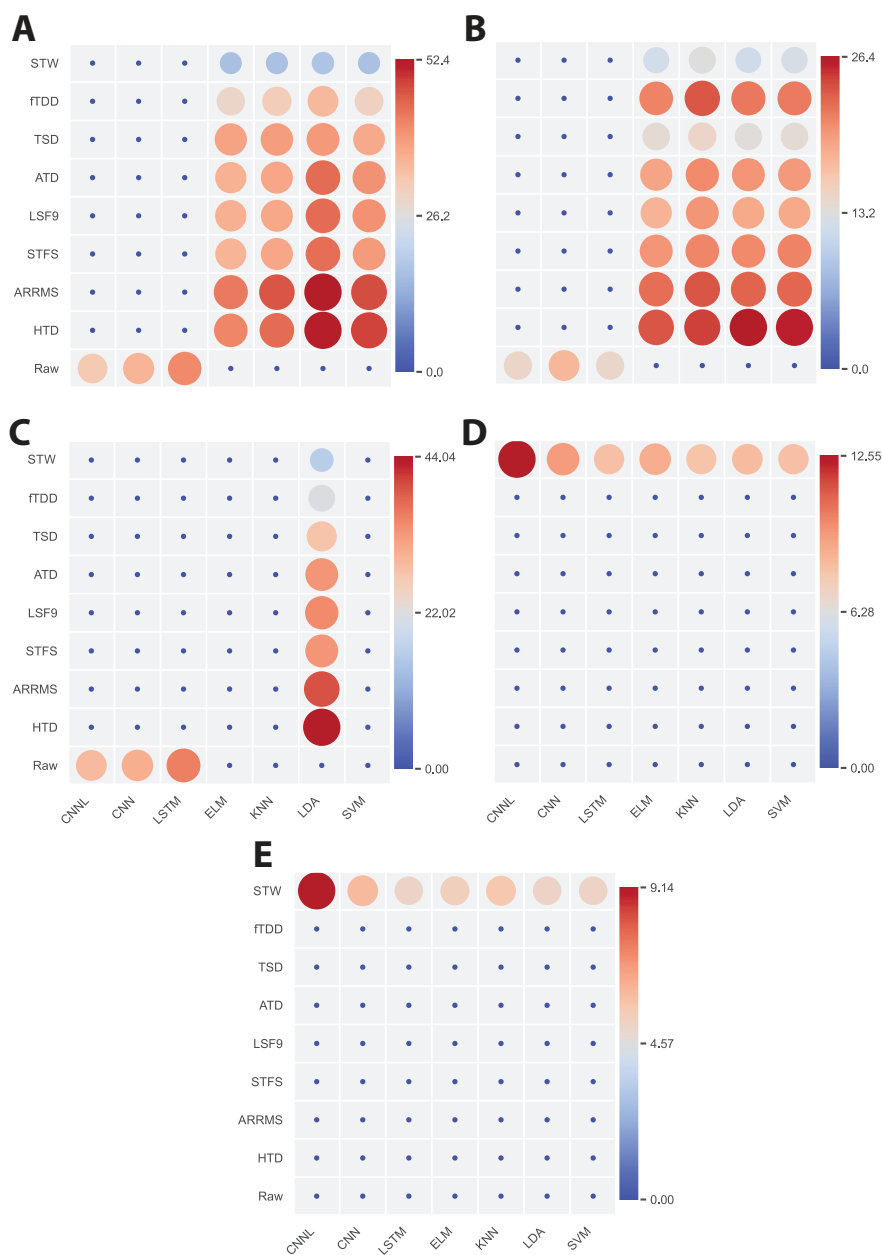
Figure 4. Circular grid plot of classification error rates averaged across all folds using different classifiers and feature extraction methods for the four data sets. Average classification error values represented by circles for DB5, SC, DB7, 3DC with MYO armband, and 3DC with 3DC armband are illustrated in A-E, respectively. The values of errors are scaled with the size and color of the circles. For each subplot, the lowest and highest error values are presented by the smallest light blue and largest dark red circles. For the cells, without circles, there was no analysis performed. CNN+LSTM is noted with CNNL, for brevity.

3.5. Class-wise Standard Deviation

Reporting average classification error or accuracy for the off-line analysis is biased and the comparison of results; achieved by a specific method, between off-line and real-
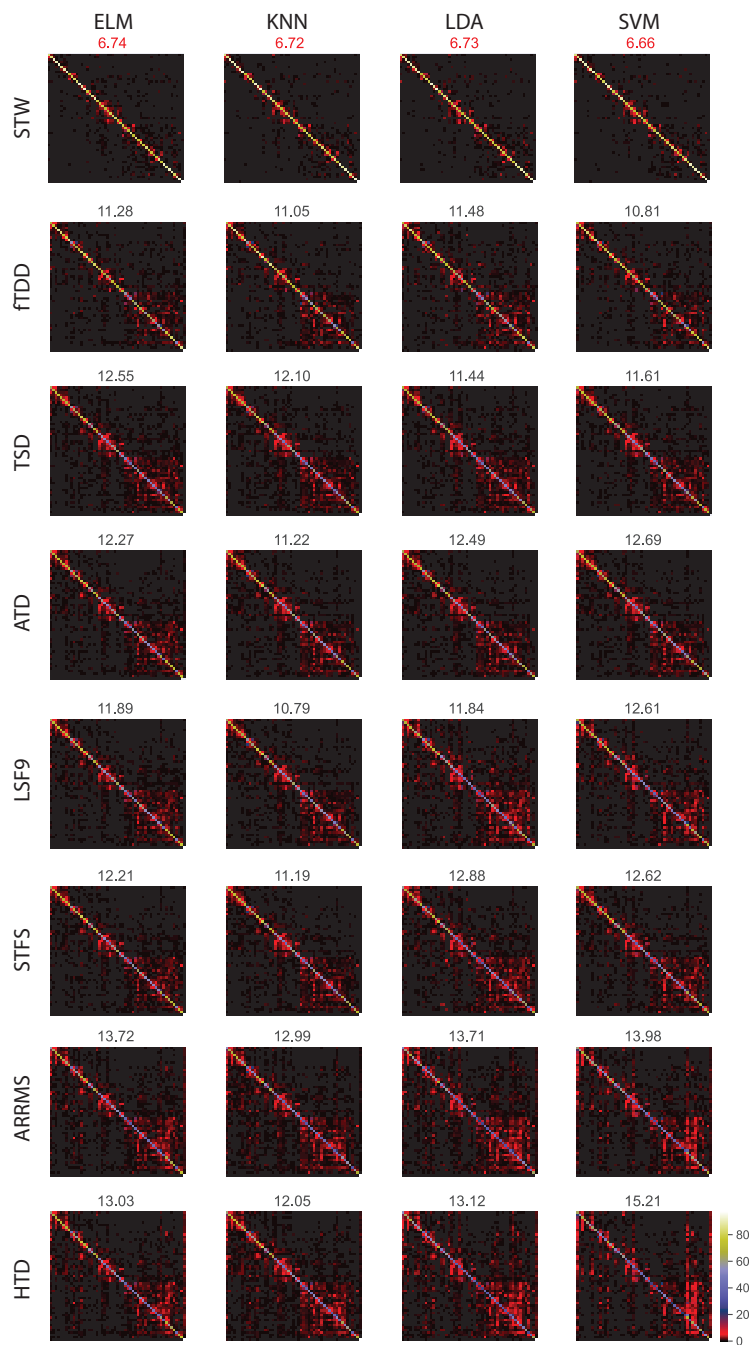
Figure 5. Average confusion matrices of the 10 participants across six folds for the DB5 dataset using different feature extraction methods as input of the four classifiers. Confusion matrices achieved by using various feature extraction algorithms are shown in different rows, respectively. The numbers above the confusion matrices indicate the CWS values achieved by feeding different feature sets into the classifiers.

time schemes will not be fair. To partially address this challenge, we used class-wise accuracy standard deviations (CWS) [67]. Averaged confusion matrices across all folds and subjects were calculated and the standard deviation of the main diagonal of the
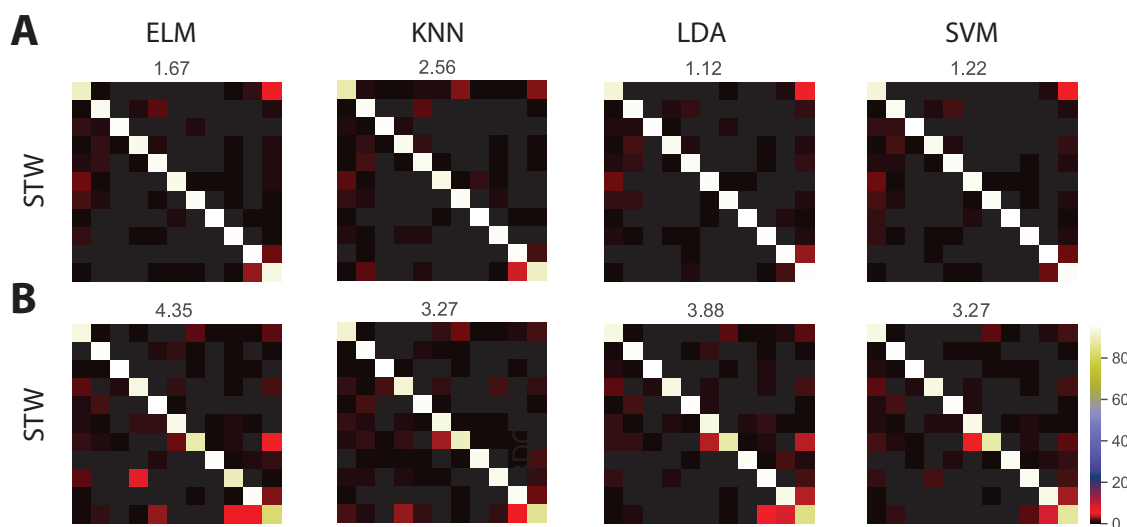
Figure 6. Average confusion matrices of the 22 participants across eight folds for the 3DC dataset using STW feature extraction method as input of the four classifiers: ELM, KNN, LDA, and SVM. Confusion matrices achieved for 3DC and MYO armbands are shown in A, and B, respectively. The numbers above the confusion matrices indicate CWS values achieved by feeding the STW feature into the classifiers.

matrices was considered as CWS. In this analysis, DB5 and 3DC were considered.

Averaged confusion matrices across 10 subjects and 6 folds using ELM, KNN, LDA, and SVM classifiers for the DB5 dataset are illustrated in figure 5. The proposed STW method could achieve the lowest CWS values among all feature sets. The results of CWS analysis for the 3DC dataset are presented in figure 6.

3.6. Windows Size

For the DB5 and 3DC, we varied the feature extraction window size from 50 msec to 250 msec. The average classification error rates across all folds and subjects for DB5, 3DC with 3DC armband, and 3DC with MYO armband are shown in figure 7(A), (B), and (C), respectively. Results show the classification errors decrease with larger windows, as expected. However, the results also show that varying the window size had a medium to large effect between consecutive windows, which could potentially indicate the robustness of the method against varying window sizes.

3.7. Computation Time

The computation time for each of the feature extraction methods, except deep learning models, was calculated on randomly generated data with 150 samples across 10 dimensions; equivalent to 150 ms with 10 channels sampled at 1000 Hz. The time required to extract the features was registered and the analysis was repeated 1000 times. Average results are displayed in Table 2.
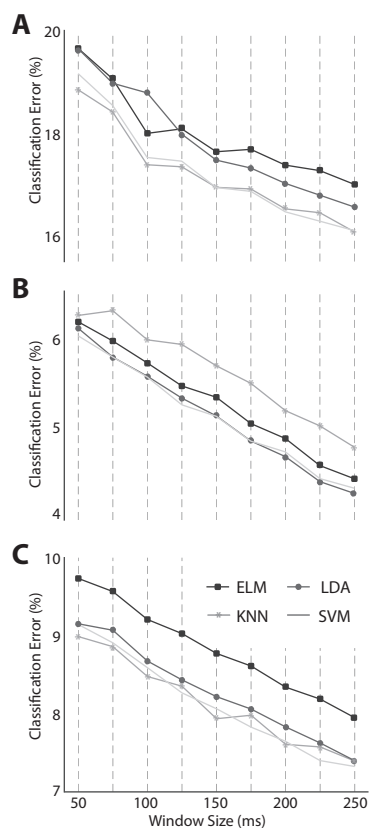
Figure 7. Average STW classification error rates with different windows sizes; 50-250 msec, achieved by ELM, LDA, KNN, and SVM classifiers. A, Average classification values for DB5 dataset. B, Average classification error rates for 3DC dataset using 3DC armband. C, Average classification error values for 3DC dataset using MYO armband.

Table 2. Time required to extract handcrafted features reported as average ± standard deviation.

Feature set	Time
LSF9	0.8053 ±0.0620 ms
TSD	0.4940 ±0.0411 ms
STFS	0.4986 ±0.0381 ms
STW	**0.4248** ±0.0632 ms
ATD	0.2115 ±0.0237 ms
HTD	0.1421 ±0.0180 ms
AR-RMS	0.1660 ±0.0509 ms
fTDD	0.0125 ±0.0354 ms

3.8. Comparison with other Studies

To evaluate the proposed STW method against recently published state-of-the-art features and classifiers a comprehensive comparison is performed on DB1 of the Ninapro database. For a fair comparison, a similar train-test approach was selected according

Table 3. Comparison of the proposed STW method with previous works on DB1 dataset.

Method	200 ms
Traditional-RF [58]	75.3%
AtzortNet [69]	66.6%
GengNet [31]	77.8%
RNN Module with raw-signal [68]	79.8%
CNN Module with raw-image1 [68]	83.5%
CNN Module with feature-signal-image1 [68]	86.3%
Hybrid CNN-RNN with raw-image1 [68]	84.7%
Hybrid CNN-RNN with feature-signal-image1 [68]	86.7%
Attention-based hybrid CNN-RNN with raw-image1 [68]	84.8%
Attention-based hybrid CNN-RNN with feature-signal-image1 [68]	87.0%
STW+SVM	**87.3**%

to the [58, 68] in such a way that repetitions 1, 3, 4, 6, 8, 9, and 10 were selected to train the classifiers, and repetitions 2, 5, and 7 were used to evaluate methods. Moreover, the length of the window was selected equal to 200 ms. Average classification accuracy achieved by STW+SVM in comparison with previous traditional features as well as deep learning models is reported in Table 3. Moreover, averaged DBI values achieved by conventional feature extraction methods and the proposed STW algorithm are illustrated in figure 4 of Supplementary Materials. Results show that STW has achieved the lowest DBI value in comparison to other methods.

4. Discussion

We presented a new hand-crafted feature extraction algorithm that borrows concepts from deep learning models and mixes these with the spatial information concept implemented by DTW. DTW was previously utilized in the literature of myoelectric control to compare training and testing templates [51–53], but it was not used to capture the spatial similarity concurrent with temporal information. The proposed STW feature was achieved by mixing spatial information with a long and short-term memory component with an attention normalization step. The STW feature outperformed the state-of-the-art feature extraction and learning algorithms such as STFS, LSF9, LSTM, CNN, and CNN+LSTM across several datasets. Our results corroborate the recent literature that the traditional features, e.g. HTD and AR-RMS, can no longer compete with new features [7, 13, 19, 23, 25, 45].

LSTM and CNN were significantly outperformed by their combination of CNN+LSTM on DB7 and DB5 which in turn demonstrates the power of the spatio-temporal feature learning. Also, CNN significantly outperformed LSTM, which in turn indicates that the significance of the spatial information captured by CNN

and CNN+LSTM. Interestingly, LSTM could only compete with HTD and AR-RMS features, which is in line with previous results from the literature suggesting that LSTM could potentially function in a better way when mixed with hand-crafted features [21]. When considering the significantly large number of models parameters for LSTM, CNN, and CNN+LSTM [32, 41, 42], STW only extracts a small number of features making it more suitable for future real-time implementations. This is further supported by the computational time requirements for STW which falls within the range of the hand-crafted features, as shown in Table 2.

The fTDD feature, with a step parameter of 15, showed a good performance on many datasets. This means that fTDD should save the extracted features from the previous 15'th window to achieve this performance, which may impact the controller delay during real-time tests. However, the design of STW allows looking back at any of the features from any window. This paper showed the simplest example of looking at the feature extracted from the preceding window only, which makes STW more attractive for real-time implementations than CNN+LSTM, CNN, LSTM, and fTDD.

Research shows how to extract optimal performance [70] and mine a trillion time series sub-sequences using DTW and the potential use of DTW for real-time problems [71, 72]. A previous study conducting 35 million experiments, on 85 datasets, with dozens of rival methods concluded that Nearest Neighbor DTW (NN-DTW) is very hard to beat. When NN-DTW can be outperformed, it is typically by a very small margin and at the cost of a huge effort in coding/complexity of implementation, and a large time and space overhead [54]. All of this make DTW an attractive algorithmic choice for our analysis in this paper. However, previous research utilizing DTW in myoelectric control only considered time-series similarity across training and testing EMG sequences or templates from individual channels [51–53]. In comparison, we utilize DTW as a spatial feature extraction step to calculate the warped similarity between multiple EMG channels and then temporally evolve the estimated similarity across time while considering long and short-term memories.

A similar processing step was also suggested in the fTDD algorithm in [19], as fTDD also fuses the features extracted from the current analysis windows (from all channels) with the same features extracted from a previous window while leaving the choice of the previous window temporal index as an empirical parameter for the user to optimize. In comparison to fTDD, we restricted the design of this stage to fuse the features extracted from the current set of windows at time step t with those coming from the nearby set of windows (1st previous, or 2nd previous, or 3rd previous, etc.) to account for the short-term memory component. To simplify the real-time design of the algorithm, we connected each module to the first previous module. This approach limited the focus of the method to short-term memory. Furthermore, a comparison of TSD and fTDD reveals that there is instability in terms of classification accuracy (figure 4). There are three main reasons for this issue. First is that DB5 and DB7 are from intact limbed people (except 2 subjects from DB7). The SC-EMG dataset includes data from five trans-radial amputees. In fact, the original paper, which proposed TSD,

showed that TSD can outperform fTDD on data from amputees. This could be related to the change of the morphology of the signals after amputation and the fact that the temporal patterns of EMG signals become harder to separate. The second reason is related to how TSD and fTDD features are extracted. fTDD focuses on the short-term temporal component only while TSD considers both temporal and spatial components. This issue motivated us to develop STW to enjoy the best of both of these methods and provide more robust performance across all databases. Moreover, TSD does not extract the inter-temporal dependencies that exist between feature extraction windows. The third reason is about the step size parameter in the fTDD algorithm, which should be optimized for data from amputees. As shown in figure 2 of Supplementary Materials, the average classification error rates are decreased with increasing the step size. This parameter plays a key role in achieving acceptable accuracy. In this feasibility study, we used a fixed step of 15. However, our proposed STW method can outperform both TSD and fTDD features for intact subjects as well as amputees with only one step.

In Figure 4, we have evaluated the proposed STW method in comparison with the conventional methods as well as deep learning models for DB5, DB7, and SC-EMG datasets. Raw EMG signals were used to evaluate the performance of the deep learning models, whereas, conventional feature sets and STW were considered as inputs for traditional classifiers. Figure 4 and Figure 1 of the Supplementary Materials show that the combination of the STW feature and traditional classifiers outperforms the combination of other feature and traditional classifiers as well as the decoding of raw EMG signals with deep learning classifiers. For the 3DC dataset, we decided to use only STW as input for both traditional classifiers and deep learning models. We aimed to investigate whether using the proposed STW feature as input for deep learning models can outperform the combination of the STW feature and traditional classifiers. For the MYO dataset, the results show that feeding an SVM with the STW feature achieves the lowest classification error rate ($8.06\pm6.22\%$). For the 3DC armband results, the combination of the STW features and the LSTM decoder resulted in the lowest error rate ($5.10\pm6.35\%$).

Attention mechanisms relating to different positions of a single sequence have become an integral part of sequence modeling and transduction models in various tasks, allowing modeling of dependencies without regard to their distance in the input or output sequences. A softmax function is usually employed for attention weights. We computed the attention component by dividing the extracted feature by their total sum to ensure that all features sum up to unity.

The major difference between the proposed method and LSTM is that LSTM learns a recurrent feature representation from the data, while STW extracts a set of hand-crafted DTW features. Additionally, STW blocks have a simpler design than LSTM and are more suitable for real-time implementations.

One possible limitation of this study is that the DTW algorithm may prove too complex for real-time implementation and may require additional hardware for parallel processing. Our results show that the time required to extract STW features is smaller

than LSF9, TSD, and STFS features. Additionally, whether using an advanced machine learning method could enhance the quality of prosthesis control in real-life settings is an open question and falls outside the scope of this study. It is worth mentioning that in this study we investigated a wide range of time-domain features and did not evaluate spectral or time-frequency features as their enhanced separability compared to the time-domain features is unknown and furthermore, it has been shown that time-domain features can outperform time-frequency features [73]. Furthermore, beyond computational complexity, we do not envisage any challenges in using the STW method for high-density EMG. As for the outliers, we did not investigate STW because the utilized databases are almost outlier-free, as they are recorded in controlled laboratory conditions.

In conclusion, we proposed a new paradigm for grasp level control of prosthetic hands. We demonstrated the feasibility of decoding EMG signals in both able-bodied and below-elbow amputees. Our algorithm warrants further investigation with real-time, user-in-the-loop experiments with people with upper-limb differences.

References

[1] Martinez IJ, Mannini A, Clemente F, Sabatini AM, Cipriani C 2020 Grasp force estimation from the transient EMG using high-density surface recordings *J. Neural Eng.* **17** 016052

[2] Tanzarella S, Muceli S, Del Vecchio A, Casolo A, Farina D 2020 Non-invasive analysis of motor neurons controlling the intrinsic and extrinsic muscles of the hand *J. Neural Eng.* **17** 046033

[3] Martínez IJ, Mannini A, Clemente F, Cipriani C 2020 Online grasp force estimation from the transient EMG *Trans. Neural Syst. Rehabil. Eng.* **28** 2333-41

[4] Chen C, Yu Y, Sheng X, Farina D, Zhu X 2021 Simultaneous and proportional control of wrist and hand movements by decoding motor unit discharges in real time *J. Neural Eng.* **18** 056010

[5] Guo W, Ma C, Wang Z, Zhang H, Farina D, Jiang N, Lin C 2021 Long exposure convolutional memory network for accurate estimation of finger kinematics from surface electromyographic signals *J. Neural Eng.* **18** 026027

[6] Xiao F, Chen Y, Zhu Y 2020 GADF/GASF-HOG: feature extraction methods for hand movement classification from surface electromyography *J. Neural Eng.* **17** 046016

[7] Khushaba RN, Scheme E, Al-Timemy AH, Phinyomark A, Al-Taee A, Al-Jumaily A 2021 A long short-term recurrent spatial-temporal fusion for myoelectric pattern recognition *Expert Syst. Appl.* **178** 114977

[8] Rodríguez-Tapia B, Soto I, Martínez DM, Arballo NC 2020 Myoelectric interfaces and related applications: current state of EMG signal processing–a systematic review *IEEE Access* **8** 7792-7805

[9] Phinyoumark A, Scheme E 2018 EMG pattern recognition in the era of big data and deep learning *Big Data and Cognitive Computing* **2** 1-21

[10] Osborn LE, Moran CW, Johannes MS, Sutton EE, Wormley JM, Dohopolski C, Nordstrom MJ, Butkus JA, Chi A, Pasquina PF, Cohen AB 2021 Extended home use of an advanced osseointegrated prosthetic arm improves function, performance, and control efficiency *J. Neural Eng.* **18** 026020

[11] Dyson M, Dupan S, Jones H, Nazarpour K 2020 Learning, generalization, and scalability of abstract myoelectric control *IEEE Trans. Neural Syst. Rehabil. Eng.* **28** 1539 - 1547

[12] Hargrove LJ, Englehart K, Hudgins B 2007 A Comparison of Surface and Intramuscular Myoelectric Signal Classification *IEEE Trans. Biomed. Eng.* **54** 847-853

[13] Krasoulis A, Vijayakumar S, Nazarpour K 2020 Multi-grip classification-based prosthesis control with two EMG-IMU sensor *IEEE Trans. Neural Sys. Rehab. Eng.* **13** 046011

[14] Kamavuako E, Scheme E, Englehart K 2016 Determination of optimum threshold values for EMG time domain features; A multi-dataset investigation *J. Neural Eng.* **28** 508-518

[15] Khushaba RN, Al-Timemy AH, Al-Ani A, Al-Jumaily A 2017 A framework of temporal-spatial descriptors-based feature extraction for improved yoelectric pattern recognition *IEEE Trans. Neural Syst. Rehabil. Eng.* **25** 1821-1831

[16] Tkach D, Huang H, Kuiken A 2010 Study of stability of time-domain features for electromyographic pattern recognition *J. Neuroeng. Rehabilitation.* **7** 1-3

[17] Khushaba RN, Al-Timemy A, Kodagoda S, Nazarpour K 2016 Combined influence of forearm orientation and muscular contraction on EMG pattern recognition *Expert Syst. Appl.* **61** 154-161

[18] Hudgins B, Parker P, Scott RN 1993 A new strategy for multifunction myoelectric control *IEEE Trans. Biomed. Eng.* **40** 82-94

[19] Khushaba RN, Al-Ani A, Al-Timemy A, Al-Jumaily A 2016 A fusion of time-domain descriptors for improved myoelectric hand control *In: Proceedings of the IEEE Symposium Series on Computational Intelligence (SSCI), Athens,* 1-6

[20] Campbell E, Phinyomark A, Scheme E 2019. Linear discriminant analysis with bayesian risk parameters for myoelectric control *In: Proceedings of the IEEE Global Conf. on Signal and Information Processing (GlobalSIP), Ottawa, ON, Canada,* 1-5

[21] Olsson A, Maleševic N, Björkman A, Antfolk C 2019 Exploiting the intertemporal structure of the

upper-limb sEMG: comparisons between an LSTM network and cross-sectional myoelectric pattern recognition methods. *In" Proceedings of the 41st Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany,* 6611-6615

[22] Phinyomark A, Khushaba RN, Scheme E 2018 Feature extraction and selection for myoelectric control based on Wearable EMG sensors *Sensors* **18** 1615

[23] Al-Taee AA, Khushaba RN, Al-Timemy AH, Al-Jumaily A 2019 Spatially filtered low-density EMG and time-domain descriptors improves hand movement recognition *In: Proceedings of the 41st Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany* 2671-2674

[24] Campbell E, Phinyomark A, Scheme E 2020 Current trends and confounding factors in myoelectric control: limb position and contraction intensity *Sensors* **20** 1613

[25] Al-Timemy AH, Khushaba RN, Bugmann G, Escudero J 2016 Improving the performance against force variation of EMG controlled multifunctional upper-limb prostheses for transradial amputees *IEEE Trans. Neural Syst. Rehabil. Eng.* **24** 650-661

[26] Phinyomark A, Quaine F, Charbonnier S, Serviere C, Tarpin-Bernard F, Laurillau Y 2013 EMG feature evaluation for improving myoelectric pattern recognition robustness *Expert Syst. Appl.* **40** 4832-4840

[27] Phinyomark A, Khushaba RN, Ibáñez-Marcelo E, Patania A, Scheme E, Petri G 2017 Navigating features: a topologically informed chart of electromyographic features space *J. R. Soc. Interface.* **14** 07-34

[28] Samadani A 2018 Gated recurrent neural networks for EMG-based hand gesture classification. a comparative study *In: Proceedings of the 40th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBC), Honolulu, HI* 1-4

[29] Simao M, Neto P, Gibaru O 2019 EMG-based online classification of gestures with recurrent neural networks *Pattern Recognit. Lett.* **128** 45-51

[30] Wang C, Guo W, Zhang H, Guo L, Huang C, Lin C 2020 sEMG-based continuous estimation of grasp movements by long-short term memory network *Biomed. Signal Process. Control* **59** 101774

[31] Geng W, Du Y, Jin W 2016 Gesture recognition by instantaneous surface EMG images *Sci. Rep.* **6** 36571

[32] Côté Allard U, Campbell E, Phinyomark A, Laviolette F, Gosselin B, Scheme E 2020 Interpreting deep learning features for myoelectric control: a comparison with handcrafted features *Front. Bioeng. Biotechnol.* **8** 158

[33] Sharma M, Malhotra H, Kumar N, Yadav J 2019 Deep learning with convolutional neural network for controlling human arm prosthesis using sEMG signals *In: Proceedings of the int Conf. on Computing, Power and Communication Technologies (GUCON), NCR New Delhi, India* 796-803

[34] Gao Q, Liu J, Ju A 2020 Hand gesture recognition using multimodal data fusion and multiscale parallel convolutional neural network for human–robot interaction *Expert Syst.* e12490

[35] Zia-Ur-Rehman M, Waris M, Gilani S, Jochumsen M, Niazi I, Jamil M, Farina D, Kamavuako E 2018 Multiday EMG-based classification of hand motions with deep learning techniques *Sensors* **18** 2497

[36] Xia P, Hu J, Peng Y 2018 EMG-Based estimation of limb movement using deep learning with recurrent convolutional neural networks *Artif. Organs.* **42** E67-77

[37] Hu Y, Wong Y, Wei W, Du Y, Kankanhalli M, Geng W 2018 A novel attention-based hybrid CNN-RNN architecture for sEMG-based gesture recognition *PLOS ONE* **13** e0206049

[38] Wu Y, Zheng B, Zhao Y 2018 Dynamic gesture recognition based on LSTM-CNN *In: Proceedings of the Chinese Automation Congress (CAC), Xi'an, China* 2446-2450.

[39] Atzori M, Muller H 2019 PaWFE: fast signal feature extraction using parallel time windows *Front. Neurorobot* **13** 74

[40] Cognolato M, Brigato L, Cid YD, Atzori M, Müller H 2019 Analyzing the trade-off between training session time and performance in myoelectric hand gesture recognition during upper limb movement *In: Proceedings of the IEEE 16th Int. Conf. Rehabilitation Robotics (ICORR), Toronto,*

ON, Canada 772-777

[41] Côté-Allard U, Gagnon-Turcotte G, Laviolette F, Gosselin B 2019 A Low-Cost, wireless, 3-D-printed custom armband for sEMG hand gesture recognition *Sensors* **19** 2811

[42] Côté Allard U, Fall CL, Drouin A, Campeau-Lecours A, Gosselin C, Glette K, Laviolette F, Gosselin B 2018 Deep learning for electromyographic hand gesture signal classification using transfer learning *IEEE Trans. Neural Syst. Rehabil. Eng.* **27** 760-771

[43] Stachaczyk M, Atashzar SF, Farina D 2020 Adaptive spatial filtering of high-density EMG for reducing the influence of noise and artefacts in myoelectric control *IEEE Trans. Neural Syst. Rehabil. Eng.* **28** 1511-1517

[44] Ebied A, Kinney-Lang E, Escudero J 2019 Consistency of muscle synergies extracted via higher-order tensor decomposition towards myoelectric control *In: Proceedings of the 9th Int. IEEE/EMBS Conf. on Neural Engineering (NER), San Francisco, CA, USA* 315-318

[45] Jabbari M, Khushaba RN, Nazarpour K 2020 EMG-based hand gesture classification with long short-term memory deep recurrent neural networks *In: Proceedings of the 42nd Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBC), Montreal, QC, Canada* 3302-3305

[46] Côté Allard U, Gagnon-Turcotte G, Phinyomark A, Glette K, Scheme E, Laviolette F, Gosselin B 2020 Unsupervised domain adversarial self-calibration for electromyographic-based gesture recognition *arXiv* 1912-11037

[47] Rabiner LR, Juang BH 1993 *Fundamentals of Speech Recognition*, Prentice Hall Signal Processing Series, USA

[48] Muller M 2007 Dynamic time warping in information retrieval for music and motion *Springer* 69-84

[49] Gavrila DM, Davis LS 1995 Towards 3-d model-based tracking and recognition of human movement: a multi-view approach *In: Proceedings of the Int. Workshop on Automatic Face-and Gesture-Recognition. IEEE Computer Society, Zurich*

[50] Schmill M, Oates T, Cohen P 1999 Learned models for continuous planning *In: Proceedings of the Seventh Int. Workshop on Artificial Intelligence and Statistics* 278-282

[51] Shuman G, Durić Z, Barbará D, Lin J, Gerber LH, 2016 Improving the recognition of grips and movements of the hand using myoelectric signals *BMC Med. Inform. Decis. Mak.* **16** 65-83

[52] Powar OS, Chemmangat K 2019 Dynamic time warping for reducing the effect of force variation on myoelectric control of hand prostheses *J. Electromyogr. Kinesiol.* **48** 152-160

[53] Powar OS, Chemmangat K 2020 Reducing the effect of wrist variation on pattern recognition of myoelectric hand prostheses control through dynamic time warping *Biomed. Signal. Process. Control.* **55** 101626

[54] Bagnall A, Lines J, Bostrom A, Large J, Koegh E 2017 The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances *Data Min. Knowl. Discov.* **31** 606–660

[55] Hahne J, Biebmann F, Jiang N, Rehbaum H, Farina D, Meinecke F, Müller KR, Parra L 2014 Linear and nonlinear regression techniques for simultaneous and proportional myoelectric control *IEEE Trans. Neural Syst. Rehabil. Eng.* **22** 269-279

[56] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A, Kaiser L, Polosukhin I 2017 Attention is all you need *In: Proceedings of the 31st Int. Conf. on Neural Information Processing Systems* 6000-6010

[57] Atzori M, Gijsberts A, Kuzborskij I, Elsig S, Hager AG, Deriaz O, Castellini C, Müller H, Caputo B 2014 Characterization of a benchmark database for myoelectric movement classification *IEEE Trans. Neural Syst. Rehabil. Eng.* **23** 73-83

[58] Atzori M, Gijsberts A, Castellini C, Caputo B, Hager AG, Elsig S, Giatsidis G, Bassetto F, Müller H 2014 Electromyography data for non-invasive naturally-controlled robotic hand prostheses *Sci. data.* **1** 1-3

[59] Pizzolato S, Tagliapietra L, Cognolato M, Reggiani M, Müller H, Atzori M 2017 Comparison of six electromyography acquisition setups on hand movement classification tasks *PLOS ONE* **12**

e0186132

[60] Krasoulis A, Kyranou I, Erden M, Nazarpour K, Vijayakumar S 2017 Improved prosthetic hand control with concurrent use of myoelectric and inertial measurements *J. Neuroeng. Rehabil.* **14** 71

[61] Scheme EJ, Englehart KB, Hudgins BS 2011 Selective classification for improved robustness of myoelectric control under nonideal conditions. *IEEE Trans. Biomed. Eng.* **58** 1698–1705

[62] Samuel OW, Asogbon M, Geng Y, Li X, Pirbhulal, S, Chen S, Naik G, Feng P, Li G 2019 Spatio-temporal based descriptor for limb movement-intent Characterization in EMG-pattern recognition system *In: Proceedings of the 41st Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany* 2637-2640

[63] Cai D, He X, Han J 2008 SRDA: An efficient algorithm for large-scale discriminant analysis *IEEE Trans. Knowl. Data. Eng.* **20** 1-12

[64] Sokolova M, Lapalme G 2009 A systematic analysis of performance measures for classification tasks *Inf. Process. Manag.* **45** 427-37

[65] Davies DL, Bouldin DW 1979 A cluster separation measure *IEEE Trans. Pattern Anal. Mach. Intell.* **2** 224-7

[66] Tallec C, Ollivier Y 2018 Can recurrent neural networks warp time? *In: Proceedings of the Int. Conf. on Learning Representations* 2018

[67] Lv, B., Sheng, X., Hao, D., Zhu, X. 2019 Relationship between offline and online metrics in myoelectric pattern recognition control based on target achievement control test *In: Proceedings 41st Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBC)* 6595-6598

[68] Hu Y, Wong Y, Wei W, Du Y, Kankanhalli M, Geng W 2018 A novel attention-based hybrid CNN-RNN architecture for sEMG-based gesture recognition *PLOS ONE* **13** e0206049

[69] Atzori M, Cognolato M, Müller H 2016 Deep learning with convolutional neural networks applied to electromyography data: A resource for the classification of movements for prosthetic hands *Front. Neurorobot.* **10** 9

[70] Mueen A, Keogh EJ 2016 Extracting optimal performance from dynamic time warping *In: Proceedings of the 22nd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining* 2129–213

[71] Rakthanmanon T, Campana B, Mueen A, Batista G, Westover B, Zhu Q, Zakaria J, Keogh E 2012 Searching and mining trillions of time series subsequences under dynamic time Warping *In: Proceedings of the 18th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining* 262-270

[72] Rakthanmanon T, Campana B, Mueen A, Batista G, Westover MB, Zhu Q, Zakaria J, Keogh E (2013) Data mining a trillion time series subsequences under dynamic time warping *In: Proceedings of the IJCAI Int. Joint Conf. on Artificial Intelligence* 3047-3051

[73] Englehart K, Hudgins B, Parker PA, Stevenson M 1999 Classification of the myoelectric signal using time-frequency based representations *Med. Eng. Phys.* **21** 431-8