Myoelectric Control with Fixed Convolution-based Time-Domain Feature Extraction: Exploring the Spatio-Temporal Interaction.

Rami N. Khushaba, *Senior Member, IEEE*, Ali H. Al-Timemy*, Oluwarotimi Williams Samuel, *Senior Member, IEEE*, and Erik J. Scheme, *Senior Member, IEEE*

***Abstract*—The role of feature extraction in electromyogram (EMG) based pattern recognition has recently been emphasized with several publications promoting deep learning (DL) solutions that outperform traditional methods. In this regard, it has been shown that the ability of DL models to extract temporal, spatial, and spatio-temporal information provides significant enhancements to the performance and generalizability of myoelectric control. Despite these advancements, it can be argued that DL models are computationally very expensive, requiring long training times, increased training data, and high computational resources, yielding solutions that may not yet be feasible for clinical translation given the available technology. The aim of this paper is, therefore, to leverage the benefits of spatio-temporal DL concepts into a computationally feasible and accurate traditional feature extraction method. Specifically, the proposed novel method extracts a set of well-known time-domain features into a matrix representation, convolves them with predetermined fixed filters, and temporally evolves the resulting features over a short and long-term basis to extract the EMG temporal dynamics. The proposed method, based on Fixed Spatio-Temporal Convolutions, offers significant reductions in the computational costs, while demonstrating a solution that can compete with, and even outperform, recent DL models. Experimental tests were performed on three databases, including one with high-density EMG (HD-EMG) signals, across a total of 44 subjects performing 22 movements, 53 movements, and 41 movements. Despite the simplification compared to deep approaches, our results show that the proposed solution significantly reduces the classification error rates by 3% to 10% in comparison to recent DL models, while being efficient for real-time implementations.**

***Index Terms*—Myoelectric signals, convolutional neural networks (CNN), feature extraction, spatio-temporal models.**

I. Introduction

PATTERN recognition (PR) algorithms have long been investigated for extracting discriminative information from Electromyogram (EMG) signals collected from human muscles, for the control of powered prosthetic limbs [1], [2]. A number of factors have been identified and reported to have substantial impact on the EMG signals' characteristics and on the performance of the EMG-PR based prosthetic control schemes. Important factors include electrodes shift, varying contraction force levels, forearm orientation and variation in limb positions, inter-subject-variability, and fatigue [2]–[6]. The fundamental discriminatory power of EMG pattern classifiers, however, has been found to largely depend on the quality of the extracted features [7], [8]. In this regard, when classifying EMG data acquired over numerous days, Phinyomark et al. [9] examined 50 feature extraction approaches for EMG pattern identification and concluded that the combination of sample Entropy, fourth order cepstrum coefficients, root mean square, and waveform length is a promising feature subset. On the other hand, the temporal descriptors of Khushaba et al. [10] further outperformed the aforementioned subset as well as other fairly complex convolutional network models [11]

Existing EMG-PR approaches can be broadly divided into two categories [2], [12]: (1) traditional methods based on feature engineering [6], [7], [9], [10], [13] and (2) Deep Learning (DL) methods based on feature learning [11], [14]–[16]. A major limitation of traditional methods has been that such feature extraction approaches generally treat the signals from the spatially distinct EMG sensors individually, that is by simply concatenating the features extracted from the individual channels to form a large feature set. However, it is generally known that human hand movements are realized by a synergistic action from multiple muscles working together, motivating algorithmic solutions that can characterize the synergistic activation patterns of different groups of muscles [17], [18]. Similarly, traditional feature extraction approaches employ a sliding window technique that, despite having some level of overlap between windows (often as much as 90%), largely ignores any long or short-term temporal dynamics of the EMG signals [13], [19].

Deep learning models like Long Short-Term Memory (LSTM) have been reported to capture the short- and long-term temporal dynamics of the collected EMG signals when applied on either the raw data or on time-domain features extracted from the EMG signals [20], [21]. While LSTM models don't necessarily leverage spatial information between channels, Convolutional Neural Networks (CNN) models excel at extracting spatial features [22], [23]. Conversely, while CNN models were not originally designed for temporal modelling and forecasting, recent works have shown that can be powerful forecasting models using temporal convolutions [24]. As a

* Corresponding author

Original paper submitted for review in July 2021.

Rami N. khushaba is with the Australian Center for Field Robotics, The University of Sydney, 8 Little Queen Street, Chippendale, 2008, NSW, Australia (e-mail: Rami.Khushaba@sydney.edu.au).

Ali H. Al-Timemy is within the Biomedical Engineering Department, Al-Khwarizmi College of Engineering, University of Baghdad, Iraq. (e-mail: ali.altimemy@kecbu.uobaghdad.edu.iq).

Oluwarotimit Williams Samuel is with the Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences. (e-mails: samuel@siat.ac.cn).

Erik J. Scheme is with the Institute of Biomedical Engineering, University of New Brunswick, Canada. (e-mail: escheme@unb.ca).

result, spatio-temporal feature extraction methods came out as a solution with traditional hand-crafted features [10], [25], hybrid deep learning models such as CNN+LSTM [26]–[28], and their combination [11], [20].

Despite the significant achievements reported in almost every publication on DL methods applied to EMG-based hand movement classification, it is vital to understand that such achievements come with substantial computational costs. Such costs are attributed to the huge amount of data required to train DL models such as CNNs and the relatively large set of parameters that need to be optimized/learned. Previous research has indicated that the number of DL models' parameters can range from anything between 34-95k in [29], 104k in [22], 200k in [11], 30-549k in [30], reaching to millions of parameters [31]. Training DL models requires allocating memory to weights, activations, gradients, data batches, and workspace, which is at least hundreds of MBs if not GBs [32]. These models are difficult to deploy especially on a resource-constrained device, let alone be trained on such resource-limited devices, which poses a significant huddle to their adoption in clinical applications. Therefore, designing a simple and efficient model that can run on resource-constrained platforms constitutes an essential aspect of machine learning driven prosthetic devices.

In this paper, we propose a new algorithmic solution that employs traditional feature extraction algorithms motivated by concepts adopted from DL models, rather than applying DL models directly. While traditional feature extraction methods are more computationally efficient than recent DL models, their performance on various databases has been challenged when compared against that of state-of-the-art DL models. Hence in this paper, we propose a mixture or hybrid of feature engineering and feature learning concepts that is formulated to combine the important characteristics of both approaches, i.e. 1) computational feasibility of traditional feature engineering with 2) the performance enhancements offered by DL models. Specifically, we adapt our time-domain based power spectrum descriptors (TD-PSD) with a convolution trick that requires no parameter optimization or learning, mixed with an attention module that is temporally evolved across the time domain. In this manner, significant reductions in the models' computational complexity can be realized compared to the direct application of DL models. At the same time, a significant performance improvement is demonstrated on several databases and, in many cases, the proposed model performs as well as or better than state of the art traditional and DL models.

II. Proposed fixed convolution-based Time Domain feature extraction (fcTDFE)

In the following sections, we describe the proposed methodology for EMG pattern recognition that was partly inspired by the operational modes of advanced DL models. The proposed algorithm is made of several building blocks that, together, form a novel extension that can be applied with any traditional hand-crafted features. This is followed by a description of the recursive long short-term memory concept borrowed from LSTM that is applied to traditional features with an attention mechanism.

A. Time-Domain Feature Extraction (TDFE)

The literature has so far presented us with a significant number of time-domain feature extraction (TDFE) algorithms that have been well investigated and utilized in myoelectric control [8], [9], [33]. In this paper, we decided to utilize the Time-Domain Power Spectrum Descriptors (TD-PSD) that was proposed in the work of Al-Timemy et al. [6] as the base features, since the TD-PSD has now been well tested and adopted in previous studies [13], [34]. The main goal here, however, is not focused on the power of these time-domain (TD) features, but on the impact of the proposed DL concepts that can be applied with any traditional feature extraction approaches to turn such TD features into something more powerful. The interested reader can apply the same approach with any other feature set. However, for the remaining of the analysis in this paper, we chose TD-PSD as the base feature set, given its simplicity and demonstrated performance across different publications.

In the TD-PSD feature extraction process, several features are derived from the raw representation of the EMG signals and their first and second derivatives. The use of derivatives is traced back to Hjorth feature extraction methods which employ the time-differentiation property of the Fourier transform to estimate the first few power-spectral moments [35]. In this regard, the time differentiation property simply states that the n^{th} derivative of a function in the time-domain, denoted as Δ^n for discrete time signals, is equivalent to multiplying the spectrum by k raised to the n^{th} power, with k being the frequency index. Based on this property, representations of the zero (m_0), second (m_2) and fourth (m_4) order moments are extracted by the TD-PSD method in addition to a sparsity measure (Sparseness), an irregularity measure (IRF), and the waveform length ratio (WLR). Assuming an EMG signal $\mathbf{x}$, with x_j being the jth sample of the signal, with j=0,1,2,3,...,$N-1$, and N is the total number of samples within the signal under investigation (window size), $X[k]$ being the complex power spectrum produced by the fast Fourier transform (FFT),$P[k]$ being the phase-excluded power spectrum (produced by FFT), i.e., the result of a multiplication of $X[k]$ by its conjugate $X^*[k]$ divided by N, ZC being the number of zero crossings, and NP being the number of peaks [35]. The equations of the above features are then described in Eq.1 to Eq.6 as shown below [6], [13]. The final set of TD-PSD features included 6 features which are $m0$, $m2$, $m4$, Sparseness, IRF, and WLR. As such, we utilized these features without modification and refer the interested reader to [6], [13] for more details on this aspect of the feature extraction process.

$$m_0 = \sqrt{\sum_{j=0}^{N-1} x[j]^2} \tag{1}$$

$$m_2 = \sqrt{\sum_{j=0}^{N-1} k^2 P[k]} = \sqrt{\frac{1}{N}\sum_{j=0}^{N-1} (kX[k])^2} = \sqrt{\frac{1}{N}\sum_{j=0}^{N-1} (\Delta x[j])^2} \tag{2}$$

$$m_4 = \sqrt{\sum_{j=0}^{N-1} k^4 P[k]} = \sqrt{\frac{1}{N}\sum_{j=0}^{N-1} (\Delta^2 x[j])^2} \quad (3)$$

$$Sparseness = \frac{m_0}{\sqrt{m_0 - m_4}\sqrt{m_0 - m_2}} \quad (4)$$

$$IRF = \log(\frac{ZC}{NP}) = \log\left(\frac{\sqrt{\frac{m_2}{m_0}}}{\sqrt{\frac{m_4}{m_2}}}\right) = \log(\frac{m_2}{\sqrt{m_0 m_4}}) \quad (5)$$

$$WLR = \log\left(\frac{\sum_{j=0}^{N-1}|\Delta x|}{\sum_{j=0}^{N-1}|\Delta^2 x|}\right) \quad (6)$$

B. Recursive Spatio-Temporal TDFE

Once each of the above feature types are extracted, additional processing steps are added to explicitly incorporate the spatial and temporal aspects. To extract spatial information from the six individual feature vectors (1 feature vector per each type of the 6 features), we converted each of the feature vectors into a 2-dimensional array (2D reshape process). Once the 2D arrays are formed, we performed a convolution process with a set of fixed filters of unit values. The main idea behind the convolution with fixed filters was to maintain the simplicity of the process, i.e., focusing on adding the feature values from the combination of channels under consideration according to the spatial filter locations. In comparison to CNN models, CNN optimization procedures continuously optimize the filters coefficients to reflect the best performance possible, while our model simply suggests fixed convolutions and relies on the addition of feature values to account for the spatial focus. This is shown schematically in Fig. 1. After extracting the fixed convolution-based features, an attention mechanism is applied. This is a process in which the relationship between the values in each two cells within each of the output arrays is calculated by a simple multiplication process that is scaled by the square root of the number of rows (or columns, as the output array is a square matrix).

To add the temporal focus components, we utilized an approach inspired by our previous work in [13]. In this approach, we fused the features extracted from the current analysis window with the features extracted from a previous window across the same dimensions. It should be noted here that such analyses can be performed by fusing with features from any previous window, however, for simplicity, we use the first previous window (and encourage the reader to investigate different window locations within the provided code). The next step towards changing the temporal focus of the proposed method involved re-configuring the method into a recurrent form, as shown in Fig. 2, to facilitate comparisons with models like LSTM. However, the combination of the short- and long-term memory components is the key feature enabling LSTM to capture the dynamics of any time-series data.

To address this paradox, we also added a long-term memory component to the design of the proposed TDFE with fixed convolutions as presented in Fig. 2 and evolve this while progressively extracting features from the collected EMG signals as shown in Fig. 3. In a similar fashion to LSTM models, the cell state component needs to be properly scaled to avoid overflow. This is due to the continuous addition of information to the cell state in our model, which LSTM models regulate with gates that dampen certain dimensions based on the data. In our proposed model, the evolution of the temporal information is handled as follows:

- First, to weigh the information coming from and to the cell state to account for longer memory within and between gestures, a weighting factor (denoted as β) was added. In this manner, the parameter β can be modified to balance the impact of the long-term component when extracting features, with larger or smaller values indicating higher or lower impact for the historical components, respectively. The equation covering the operation of TDFE is given by

$$TDFE = fcTDFE_t \odot fcTDFE_{t-1} + \beta \times CellState \quad (7)$$

- After each time step, a scaling operation is performed to tune down the values within the cell state by dividing the corresponding time step (t). In the next time step, the cell state factor is multiplied by $t-1$, added to the fcTDFE features from the current time steps, and scaled down again by t. A clipping process can also be applied here to ensure the values will not grow beyond a threshold. However, the progressive division by the time step index guarantees to shrink down the values after the first 20 to 30 time steps. In most cases, this is a sufficient period to help figure out the long-term memory of the movement, as our experiments indicated empirically across several subjects and datasets.

III. MATERIALS AND METHODS

A. Description of the EMG Datasets.

To promote the development of robust EMG pattern recognition algorithms, cross-validation across multiple datasets has been recommended [2]. In this regard, three different EMG databases were selected and used to assess the performance of the proposed fcTDFE method in comparison with several other state-of-the-art EMG feature extraction methods from recent literature. The details of the utilized datasets are given in Table. I.

TABLE I: Details of the utilized EMG datasets.

Dataset	Subjects	Channels	Samp.Freq (Hz)	Classes	Placement
Ninapro DB5 [36]	10	16 (2 MYOs)	200	53	Forearm
Ninapro DB7 [37]	22	12	2000	41	Forearm, biceps & triceps
HD-EMG TBI [38]	12	56	1000	22	Forearm & hand

The industry benchmark Ninapro repository was selected as the first source of EMG databases. Specifically, database-7 (DB7) [37] (20 intact-limbed, 2 amputees, 40 movements in addition to rest) and database-5 (DB5) [36] (10 intact-limbed subjects, 52 different movements in addition to rest), were considered from this repository. Different numbers of

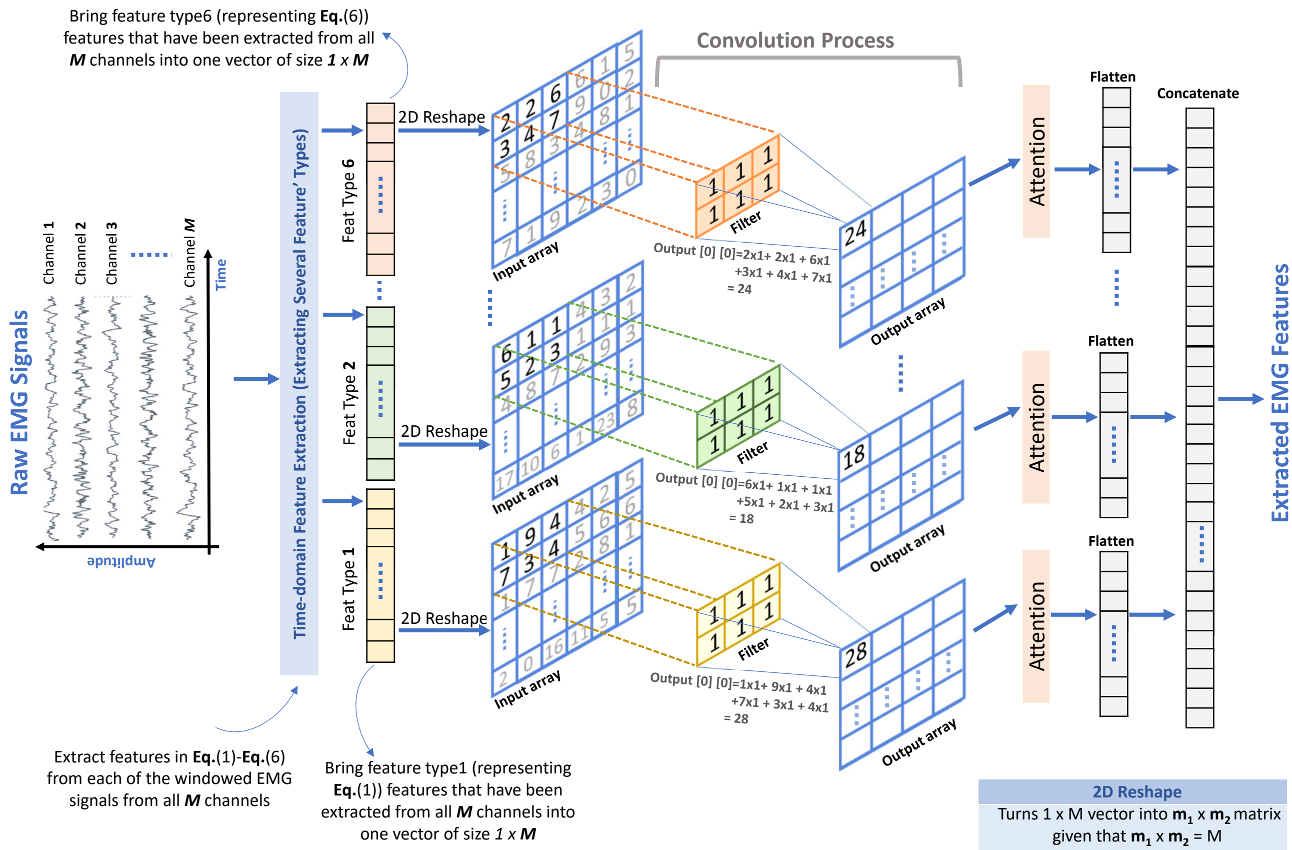

Fig. 1: Proposed fixed-convolution based time-domain feature extraction (fcTDFE). A hypothetical example is shown with a convolution filter size of 2×3, although this is a parameter to be set by the user just like in a typical CNN model. Different features are extracted from the individual signals, forming vectors of individual feature types. These vectors are reshaped into 2D matrices to allow subsequent convolution processes.

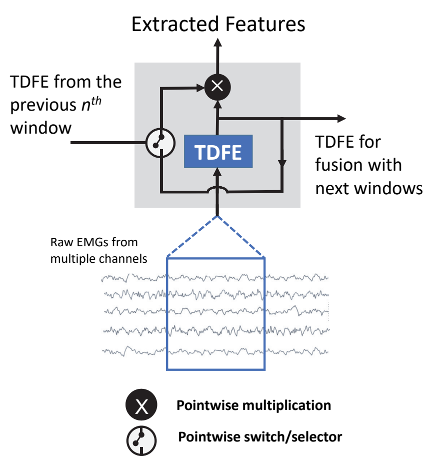

Fig. 2: A recurrent version of our proposed fcTDFE with short-term memory focus.

subjects were included in each of these databases, with each movement being repeated 6 times by each of the subjects (a few seconds of rest in-between repeats). Two wearable MYO armbands placed on the forearm were considered in DB5, with the benefits associated with this armband including its small size and low cost. It is important to mention here that the low sampling frequency of 200 Hz fails to capture large components of the EMG spectrum, which in turns constitutes a more challenging classification task [31], [36]. A total of 16 EMG channels (8 channels per MYO unit) were hence utilized from the two armbands, with one armband next to the other as shown in Fig. 4. Due to the significant loss in higher-frequency content of the EMG signals provided by the MYO armbands, it has been reported that no number of existing features can be combined to remedy or compensate for such loss. Hence, testing the effectiveness of any newly proposed features on such armbands is desirable [7].

On the other hand, DB7 (the other dataset from the Ninapro repository) was collected with high quality laboratory hardware following the Ninapro protocol. It was used to evaluate the performance of the algorithms in this study under various data sampling conditions (2000 Hz for DB7 vs. 200 Hz for MYO) and participants (amputees vs. intact-limbed). In terms of the testing scheme, a leave-one-trial-out (LOTO) scheme was employed. To limit the risk of over-fitting, 5 trials were kept for training and one trial was preserved for testing in each fold. The results of all folds were then averaged together to obtain the final result.

The final database (TBI) comprises high density EMG (HD-EMG) [38] recorded from 12 patients with traumatic brain

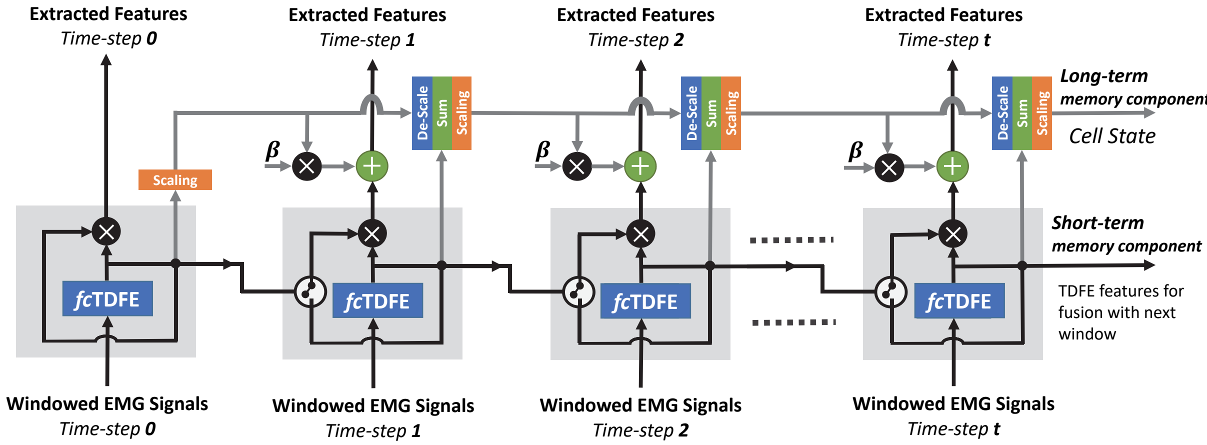

Fig. 3: Block diagram of the proposed long short-term memory time-domain feature extraction with fixed convolutions.

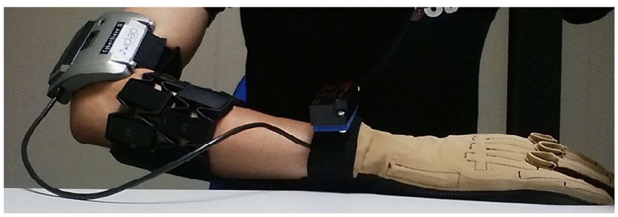

Fig. 4: Double MYO setup in Ninapro DB5 (image adopted from [36], under the Creative Commons Attribution License)

injury (TBI), and each subject was asked to perform 21 different forearm and hand movements in addition to a no movement task.

B. Details of the EMG PR System

An overlapped windowing scheme was utilized during the feature extraction process, with window sizes of 150 ms and window increments of 50 ms. For comparison, these windowing parameters were used to extract several existing and notable features reported in the literature in addition to the proposed fcTDFE. Specifically, the following feature extraction methods were investigated:

- **HTD**: Hudgins' TD feature set defined in [39]: Mean Absolute Value (MAV), MAV slope (MAVS), Waveform length (WL), Slope Sign Changes (SSC), and ZC;
- **AR-RMS**: made of a combination of the RMS and the 6th-order AR model parameters;
- **LSF9**: originally defined in [7], this set is made up of ZC, RMS, L-scale, Mean Value of the Square Root, Maximum Fractal Length, Willison Amplitude, Integrated Absolute Value, Variance, and Difference Absolute Standard Deviation Value;
- **ATD**: Combined TD and AR as defined in [37], made up of WL, MAV, Log-Variance (LogVar) and 4th-order AR model parameters,
- **STFS**: Spatio-temporal features from [25], made up of Normalized Root-Square Coefficient of 1st and 2nd differential derivatives, Integral Square Descriptor, an estimate of Mean Derivative of the higher order moments per sliding window, Mean Log Kernel, and a measure of Spatial Muscle Information.
- **TD-PSD** [6]: Six time domain based power spectrum features representing the first 3 even power spectrum Moments, a Sparsity Measure, with an Irregularity Factor and WL ratio.
- **fTDD**[1] [13]: The fusion of the TD-PSD features extracted from each analysis window with the same features extracted from a previous window that is *nSteps* away. The steps parameter of this method was empirically chosen as 15. For this analysis, we used the new version proposed by the original authors from the GitHub repository cited in the footnote below.
- **TSD**[2] [10]**:** The temporal spatial-descriptors feature extraction approach.

As the various feature extraction methods provided different numbers of features, a dimensionality reduction method was utilized to account for different feature set sizes and ensure a more fair comparison. Specifically, the Spectral Regression (SR) feature projection method [40] was used to reduce the dimensionality of all feature sets to c-1, with c being the number of classes in the corresponding dataset.

In terms of the classification models, the following traditional classifiers were evaluated, given their common use in the literature: Linear Discriminant Analysis (LDA), Support Vector Machines (SVM), Extreme Learning Machine (ELM), and K-Nearest Neighbour (KNN). As the performance of SVM is sensitive to the regularization parameter C and the kernel function parameter γ, therefore for each dataset, the SVM parameters were optimized . For KNN, the parameter K was empirically set to k=5, and for the ELM classifier, one hidden layer was used with 1250 neurons.

In order to verify the statistical significance of the achieved classification results, the Wilcoxon signed rank test was applied, with the results being considered significant for a p-value < 0.05. Finally, Cohen's effect size, d, was utilized to measure the size (or meaningfulness) of differences observed. For paired samples, this is defined as the difference between

[1]https://github.com/RamiKhushaba/getfTDDfeatures
[2]https://github.com/RamiKhushaba/getTSDfeat

two group means divided by the standard deviation [7]. Here, an effect size of 0.2 represents a moderate effect, 0.5 represents a medium effect, 0.8 represents a large effect, 1.2 represents a very large effect, while greater than 2.0 represents a gigantic influence. The direction of the effect was calculated by subtracting the mean of y from the mean of x to calculate Cohen's d value for testing x vs. y.

To investigate the performance of the proposed fcTDFE against DL methods from the literature, we compared fcTDFE against 3 methods; 1) LSTM, 2) CNN, and 3) CNN+LSTM. Fig. 5 illustrates the architectures of these three models. It is noteworthy that there is likely one best architecture of network layers for each and every scenario, and therefore extensive empirical testing is needed for each database to select the architecture that achieves the optimal performance. Nevertheless, because of their high capacity, this approach may likely result in over-tuning of hyper-parameters, limiting their generalizability. The deep learning model architectures used here are therefore reflective of those presented by the state-of-the-art [20], [29], yielding higher confidence in their validity. The raw EMG was used as inputs to these models, as recent experimental evidence suggests that raw EMG inputs to models such as CNN (or even CNN+LSTM) may be more useful than Spectrogram images [29]. Therefore, the Root Mean Square (RMS) value of the raw EMG signals was employed, generating NC scalar values for each of the 150 ms analysis windows (with NC being the number of EMG channels in each dataset). After that, we multiplied each generated vector of size ($NC \times 1$) by its transpose ($1 \times NC$), to turn the RMS values into pseudo-images, to produce images of size ($NC \times NC$). The final step was to logarithmically scale the resultant images, which were then provided as inputs to the CNN and the combination of CNN and LSTM models. For the LSTM model, the raw samples of the EMG signal of each of the 150 ms windows of the NC channels were provided and utilized to evaluate its capability for learning an effective representation of the features.

All experiments were conducted in MATLAB 2019b on a computer with an i7 core processor, an NVIDIA GeForce RTX 2060 GPU, and 16 GB of RAM. When assessing the proposed fcTDFE vs. all other models, the same trial allocations were selected for training/testing .

Two versions of the proposed fcTDFE were employed to investigate the impact of β from Eq.7 on its performance; the first, denoted as fcTDFE1 with a $\beta = 0.75$ and the second, denoted as fcTDFE2 with $\beta = 2.0$. Larger values of β did not achieve a statistically significant impact on the performance of fcTDFE and hence the use of the aforementioned parameters.

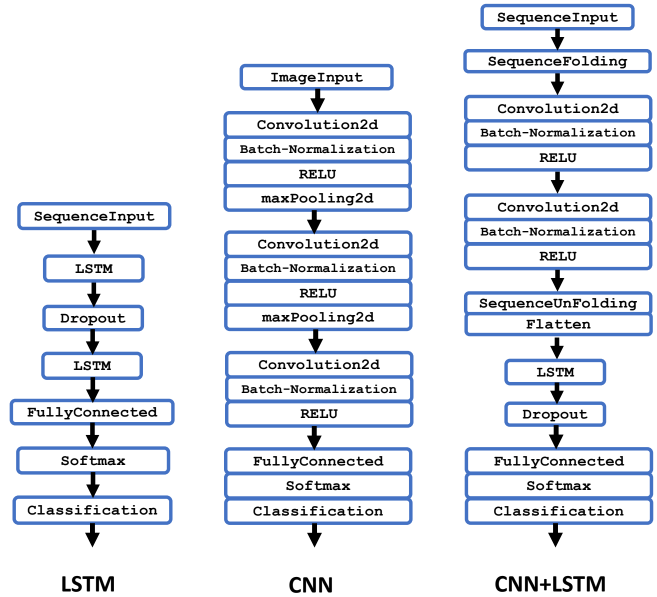

Fig. 5: The architectures of the comparison DL models.

IV. Results

A. Results of DB5

The average classification results of applying all the aforementioned feature extraction methods, in addition to the proposed fcTDFE, with all classifiers are shown in Fig.6 as a box plot. In terms of the utilized deep learning methods, the combination of CNN+LSTM performed significantly better than the individual model (CNN or LSTM) ($p < 0.01$ and $d > 2$ for both tests). The combined CNN+LSTM also performed significantly better than all other feature extraction methods (with all classifiers), except fTDD while using KNN classifier and any version of the proposed fcTDFE. Both values of β, with any of the classifiers significantly outperformed CNN+LSTM ($p < 0.01$ and $d > 2$ for fcTDFE against CNN+LSTM) and all other methods ($p < 0.01$). It is also interesting to see that the difference between the proposed fcTDFE and the original TD-PSD, upon which fcTDFE was developed, is more than 20$\%$ across all of the utilized classifiers. This provides compelling evidence of the impact of newly proposed fixed temporal convolutions with the long and short-term components. In terms of the remaining traditional feature extraction methods, it can be seen that the performance of most of the methods like HTD, AR-RMS, ATD, LSF9, and TD-PSD with different classifiers were not significantly different from that of the LSTM and CNN. It can be also noticed that most methods had some variability in the achieved error rates with the different classifiers, while this variability was much lower for TSD and the proposed fcTDFE. For the two versions of fcTDFE, no major differences were observed between the classifiers for fcTDFE2, while fcTDFE1 had a slightly higher variability.

As the results presented in Fig.6 represented the average of the LOTO testing scheme, we opted to investigate the performance of the proposed method while leaving each of the trials out for testing. The results in in Fig.7 show a significant decrease in the average classification errors across all subjects while moving away from the first trial. This can be justified by the fact that the subjects may have not been sufficiently trained on the experimental procedure before beginning, but converged after a few trials (until possibly fatiguing by the sixth trial).

Finally, the results in Fig.6 also showed that fcTDFE2 performed significantly better than fcTDFE1, which in this case indicated the importance of selecting the right value for β parameter. Hence, we employ fcTDFE with $\beta = 2.0$, for the subsequent analyses.

B. Results of DB7

To focus on the effect of features (and given similar classifier trends as seen in DB5), the average classification error rates across DB7 are shown in Fig.8 using the LDA classifier. The LDA classifier has been heavily investigated in the literature for EMG pattern recognition and hence the choice of LDA for this experiment, as both DB5 and DB7 were recorded following the same procedures. As in DB5, CNN+LSTM significantly outperformed the individual models of CNN and LSTM ($p < 0.01$, $d > 2$), while CNN also outperforming LSTM ($p < 0.01$, $d > 2$). While the performance of TSD (that is, LDA classification of the TSD features) and CNN+LSTM appear similar in the figure, TSD significantly outperformed CNN+LSTM ($p < 0.01$ and $d = 0.7154$). Although, fTDD significantly outperformed all remaining traditional methods, the proposed fcTDFE significantly outperformed it and all other tested methods ($p < 0.01$ and $d > 2$ for all tests).

The trial based classification results presented in Fig.9 showed a similar trend to those in DB5, with the classification errors when leaving the first trial out being much higher than those of the remaining trials. This was consistent across both intact-limbed and amputee subjects, and should be considered in future studies using these datasets.

C. Results of HD-EMG TBI dataset

Previous research has indicated that the interaction between the spatial and temporal information may enable the use of smaller window sizes (which may have benefits in terms of responsiveness) [41]. Hence, to scale up the challenge for all methods, windows sizes of 32 ms were utilized with windows increments of 15 ms. The average classification results for these experiments are shown in Fig.10. Contrary to the previous results on DB5 and DB7, TSD significantly outperformed fTDD. Yet still, fcTDFE outperformed all other methods ($p < 0.01$ with $d > 2$ for all tests, except for the comparison against TSD giving $d = 0.63$). The combination of CNN+LSTM again outperformed the individual CNN and LSTM models, but with no significant differences compared to fTDD/LDA and fTDD/SVM, while fTDD/ELM and fTDD/KNN showed minimal differences in terms of statistical significance ($p = 0.04$). This is, yet again, further evidence that deep learning models like LSTM, CNN and even CNN+LSTM are not necessarily better than carefully designed traditional feature extraction methods. Also, this may especially be the case for HD-EMG setups with sufficient spatial resolution.

D. Computation Time

In this part of the analysis, we benchmarked the computational time for all utilized feature extraction and learning methods in comparison to the proposed fcTDFE. It should be noted here that the training times for the different DL models were excluded from this part of the analysis as such models are known to be computationally demanding, and they are iterative methods that require a pre-specified number of epochs to train such models (unlike the analytical-based methods of traditional feature extraction methods). A randomly generated dataset with 150 samples across 10 dimensions was utilized, which is equivalent to 150 ms with 10 channels when sampled at a rate of 1000 Hz. The time required to extract the different features was then calculated in MATLAB as the analysis was repeated 1000 times for each feature type and the time was averaged, with the results displayed in Table II. The average testing time per window consumed by the DL models is also shown. This is the average time taken by the DL models to pass 150 msec worth of data from 10 channels through all the deep layers of the utilized LSTM, CNN, and LSTM+CNN models.

TABLE II: The Computational time of fcTDFE as well as different feature extraction methods.

Feature set	Time (ms)
HTD	0.2636 ms
AR-RMS	0.3780 ms
ATD	0.4822 ms
fTDD	0.1114 ms
LSF9	0.7878 ms
TD-PSD	0.1902 ms
STFS	0.6691 ms
TSD	0.8523 ms
fcTDFE	0.8536 ms
LSTM	3.8884 ms
CNN	3.2824 ms
CNN+LSTM	7.5730 ms

V. Discussion

The results of this paper demonstrated that leveraging concepts from deep learning can be used to design extremely powerful feature extraction techniques. They also highlight that deep learning models such as CNN plus LSTM, CNN, and LSTM do not necessarily outperform traditional feature engineering based methods. This has been demonstrated in this research across several databases with a varying number of subjects, hand movements, sampling frequencies and EMG electrodes. Although they are able to learn new features, they do so without context or the benefit of a priori information garnered across decades of research. Importantly, their ability to do so on each new user/dataset is of little added value in this application, and may in fact lead to reduced generalization. Nevertheless, the combination of CNN+LSTM always outperformed the individual models of CNN and LSTM, reinforcing that the powerful aspects of these models lies in their combined ability to capture spatial and temporal information. Similarly, the fTDD model proposed by Khushaba et. al. [13] showed that it can easily challenge the performance of CNN+LSTM by employing time-domain features with a temporal trick that links the information extracted from each current analysis window to that of a previous window (similar

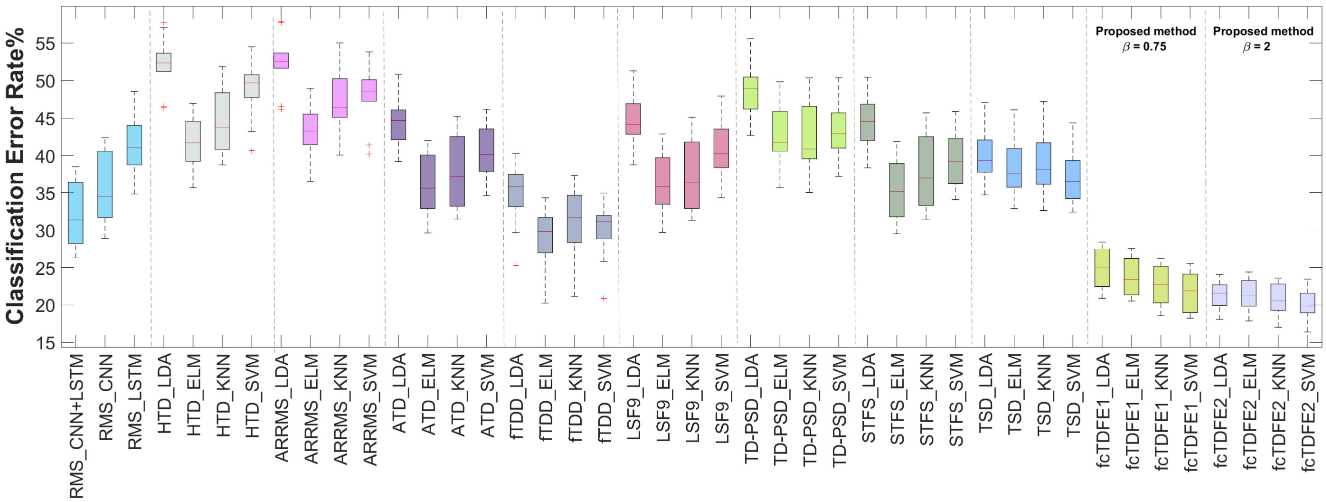

Fig. 6: DB5 classification results using all different feature extraction and classification methods.

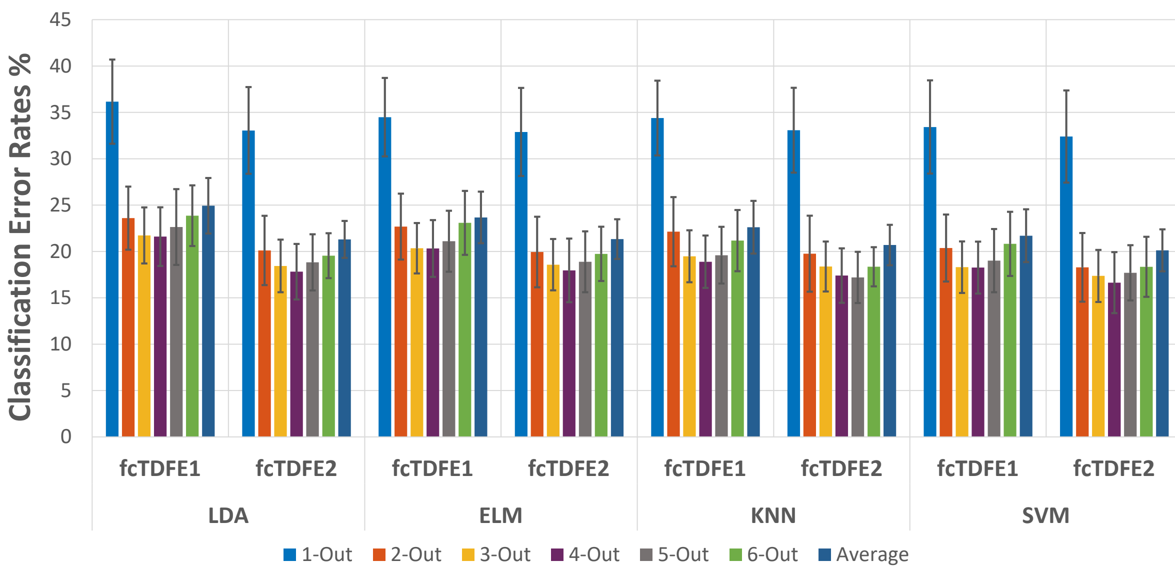

Fig. 7: DB5 average classification results per each trial with the LOTO testing scheme across the four different classifiers with the proposed fcTDFE method. Bars represent standard deviations.

to the short-term memory focus of LSTM). More importantly, the newly proposed version of fTDD that includes a focus on the inter-temporal relationship between the features extracted from the different windows clearly enhanced the performance of fTDD in relation to all other traditional methods.

Interestingly, the temporal-spatial feature engineering method TSD, presented in [10], was able to compete with fTDD on some databases, indicating that the temporal and spatial information extraction aspects can act differently on different databases, motivating the combination of these components in one method like we did with the proposed fcTDFE. The proposed fcTDFE further added upon the performance of fTDD by considering the spatial component of the feature extraction process, allowing fcTDFE to outperform all other methods.

While using very small windows sizes and increments with the TBI database, which may increase the variance of the extracted features [42], a large number of samples were extracted which theoretically should provide the deep learning models with enough data to train the corresponding models to discriminate between the underlying patterns. However, the classification error results clearly indicated that traditional methods can easily challenge the CNN+LSTM, CNN and LSTM models. Additionally, all of the fcTDFE, TSD and fTDD algorithms achieved average error rates which were below the 10% threshold previously suggested as an acceptable error for a usable EMG pattern recognition system [43].

Across the various comparisons between features, classifiers, and datasets, the proposed fcTDFE method outperformed all other approaches by a considerable margin, whether deep

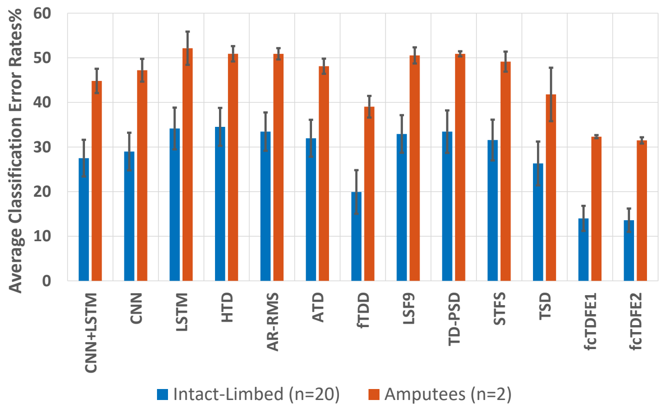

Fig. 8: DB7 average classification results using all different feature extraction methods with the LDA classifier. Bars represent standard deviations.

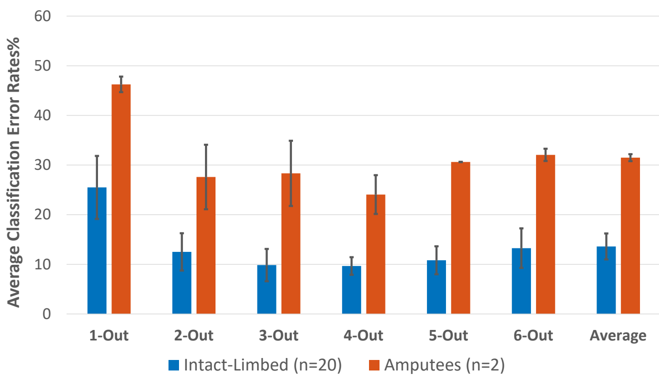

Fig. 9: DB7 Average trial based classification results using fcTDFE with LDA classifier, that is when leaving each of the trails out for testing and training based on the remaining trials.

learning based and traditional feature engineering methods. Despite this, all of the traditional feature engineering based methods, including fcTDFE, required less than 1 ms to extract across an average of 10 channels of 150 ms windows (computed across 1000 runs), which is easily within the accepted frame increments typically used for continuous control. The relative increase in computation time for TSD and the proposed fcTDFE relative to other features is due to the spatial consideration of the relationship between the different feature combinations, i.e., the attention mechanism. Nevertheless, all of these methods compare favourably with DL models because of their iterative nature during training, and computationally demanding structures. Perhaps as importantly, whereas the DL models require very large numbers of parameters to be computed and stored, especially the combined CNN+LSTM, the proposed fcTDFE generated at most a few hundred features which further reduces the computational requirements for an online application.

In our ongoing research, we aim to verify the suitability of different and larger sets of traditional features within this framework of fixed convolutions with long and short-term memory components. We also aim to investigate the usefulness of this approach for continuous estimation of hand movements by estimating hand joint angles which could be a difficult task because of the varying importance of different joints in the grasp movements [44]. Currently, we are further developing our fcTDFE method for real-time control, by investigating other performance measures as the classification accuracy alone may not to be the best for evaluating the performance of real time systems. In this direction, several offline metrics have recently shown promise as better predictors for the usability metric throughput including feature efficiency, mean semi-principal axes, and mean absolute value [45]. Our research in this direction continues as we target how to properly select the right metrics to measure the real-time performance while using our proposed fcTDFE.

Finally, the codes of the proposed fcTDFE method will be made available online[3] to help advance the research in this field.

VI. Conclusion

In this work, a time-domain feature extraction approach based on spatio-temporal convolutions was proposed to improve the performance of existing myoelectric control systems. Two versions of the proposed fcTDFE were implemented, fcTDFE1 and fcTDFE2, with two weighting factors (β), and tested across three benchmark EMG databases including HD-EMG for TBI patients. The performance of the proposed fcTDFE method significantly outperformed the state-of-the-art conventional feature extraction methods, with significant reductions in the classification errors. Significant reduction in classification error (3%-10%) were also observed when compared against deep learning methods, i.e. CNN, LSTM and CNN+LSTM. Furthermore, despite using one of the shortest windows reported in the literature for the HD-EMG TBI dataset, only 32 ms with 15 ms increments, our proposed fcTDFE method outperformed all methods including deep learning, approaching classification performances similar to those of usable EMG-based pattern recognition systems. Importantly, the proposed fcTDFE had a low computational cost of less than 1 ms and a simple time-domain implementation suggesting that it will be suitable for real time implementation in future research. The proposed method is a meaningful step forward towards improving myoelectric control, benefiting from the spatio-temporal convolutions despite having low computation cost.

Acknowledgment

The work of the third author was supported by the National Natural Science Foundation of China (82050410452) and Shenzhen Governmental Basic Research Grant (JCYJ20180507182508857). Also, we like to thank Dr. G. Yanjuan for being instrumental in providing part of the dataset used for analysis in the work.

References

[1] B. Rodríguez-Tapia, I. Soto, D. M. Martínez and N. C. Arballo, “Myoelectric Interfaces and Related Applications: Current State of EMG Signal Processing–A Systematic Review,”. *IEEE Access*, vol. 8, pp. 7792–7805, 2020, doi: 10.1109/ACCESS.2019.2963881.

[3]https://github.com/RamiKhushaba/

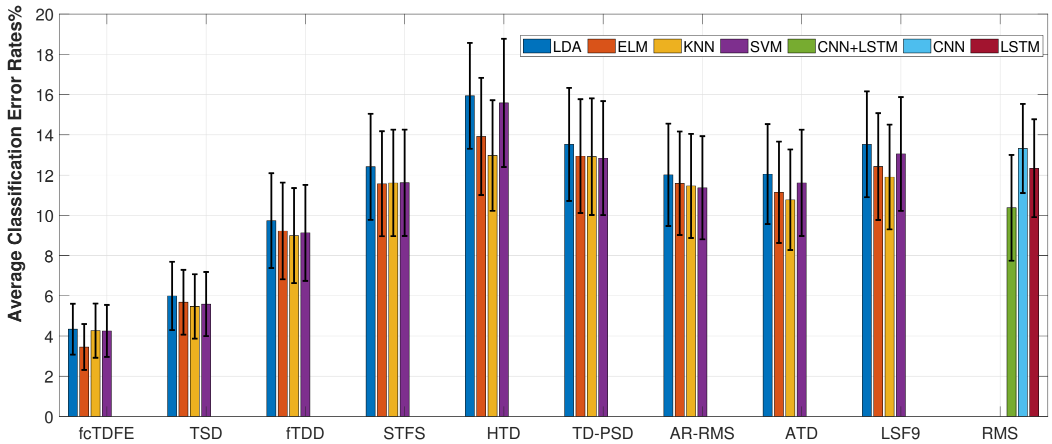

Fig. 10: Average classification error results on the HD-EMG TBI database. Bars represent standard error.

[2] A. Phinyomark and E. Scheme, "EMG Pattern Recognition in the Era of Big Data and Deep Learning". *Big Data and Cognitive Computing*. vol. 2, no. 21, 2018, 10.3390/bdcc2030021.

[3] R. N. Khushaba, A. Al-Timemy, S. Kodagoda, and K. Nazarpour, "Combined Influence of Forearm Orientation and Muscular Contraction on EMG Pattern Recognition", Expert Systems with Applications, vol 61, no.1, pp.154-161, 2016,

[4] W. Guo, X. Sheng, H. Liu and X. Zhu, "Toward an Enhanced Human–Machine Interface for Upper-Limb Prosthesis Control With Combined EMG and NIRS Signals," in IEEE Transactions on Human-Machine Systems, vol. 47, no. 4, pp. 564-575, Aug. 2017, doi: 10.1109/THMS.2016.2641389.

[5] J. -i. Furukawa, S. Chiyohara, T. Teramae, A. Takai and J. Morimoto, "A Collaborative Filtering Approach Toward Plug-and-Play Myoelectric Robot Control," in IEEE Transactions on Human-Machine Systems, 2021, doi: 10.1109/THMS.2021.3098115.

[6] A. H. Al-Timemy, R. N. Khushaba, G. Bugmann and J. Escudero, "Improving the Performance Against Force Variation of EMG Controlled Multifunctional Upper-Limb Prostheses for Transradial Amputees," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 24, no. 6, pp. 650-661, June 2016, doi: 10.1109/TNSRE.2015.2445634.

[7] A. Phinyomark, R. N. Khushaba, and E. Scheme, "Feature Extraction and Selection for Myoelectric Control Based on Wearable EMG Sensors". Sensors (Basel). vol. 18, no. 5, 2018, doi:10.3390/s18051615.

[8] A. Phinyomark, R. N. Khushaba, A. Patania, E. Ibáñez-Marcelo, E. Scheme, and G. Petri, "Navigating Features: A Topologically-Informed Chart of EMG Features Space", Journal of the Royal Society Interface (RSIF), vol. 14, no. 137, pp. 1-12, 2017.

[9] A. Phinyomark, F. Quaine, S. Charbonnier, C. Serviere, F. Tarpin-Bernard, Y. Laurillau, "EMG feature evaluation for improving myoelectric pattern recognition robustness", Expert Systems with Applications, vol. 40, no. 12, pp. 4832–4840, 2013.

[10] R. N. Khushaba, A. H. Al-Timemy, A. Al-Ani and A. Al-Jumaily, "A Framework of Temporal-Spatial Descriptors-Based Feature Extraction for Improved Myoelectric Pattern Recognition," in IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 25, no. 10, pp. 1821-1831, Oct. 2017, doi: 10.1109/TNSRE.2017.2687520.

[11] U. Côté-Allard et al., "Unsupervised Domain Adversarial Self-Calibration for Electromyography-Based Gesture Recognition," in IEEE Access, vol. 8, pp. 177941-177955, 2020, doi: 10.1109/ACCESS.2020.3027497.

[12] R. N. Khushaba, A. Phinyomark, A. H. Al-Timemy and E. Scheme, "Recursive Multi-Signal Temporal Fusions With Attention Mechanism Improves EMG Feature Extraction," in IEEE Transactions on Artificial Intelligence, vol. 1, no. 2, pp. 139-150, Oct. 2020, doi: 10.1109/TAI.2020.3046160.

[13] R. N. Khushaba, A. Al-Ani, A. Al-Timemy and A. Al-Jumaily, "A fusion of time-domain descriptors for improved myoelectric hand control", IEEE Symposium Series on Computational Intelligence (SSCI), Athens, 2016, pp. 1-6, doi: 10.1109/SSCI.2016.7850064.

[14] E. Rahimian, S. Zabihi, A. Asif, D. Farina, S. F. Atashzar and A. Mohammadi, "FS-HGR: Few-shot Learning for Hand Gesture Recognition via ElectroMyography," in IEEE Transactions on Neural Systems and Rehabilitation Engineering, doi: 10.1109/TNSRE.2021.3077413.

[15] Z. Li, J. Zuo, Z. Han, X. Han, C. Sun and Z. Wang, "Intelligent Classification of Multi-Gesture EMG Signals Based on LSTM," 2020 International Conference on Artificial Intelligence and Electromechanical Automation (AIEA), 2020, pp. 62-65, doi: 10.1109/AIEA51086.2020.00020.

[16] A. Ameri, M. A. Akhaee, E. Scheme and K. Englehart, "A Deep Transfer Learning Approach to Reducing the Effect of Electrode Shift in EMG Pattern Recognition-Based Control," in IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 28, no. 2, pp. 370-379, Feb. 2020, doi: 10.1109/TNSRE.2019.2962189.

[17] A. A. Al Taee, R. N. Khushaba and A. Al-Jumaily, "Spatially Filtered Low-Density EMG and Time-Domain Descriptors Improves Hand Movement Recognition," 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2019, pp. 2671-2674, doi: 10.1109/EMBC.2019.8857289.

[18] J. Ma, N. V. Thakor and F. Matsuno, "Hand and Wrist Movement Control of Myoelectric Prosthesis Based on Synergy," in IEEE Transactions on Human-Machine Systems, vol. 45, no. 1, pp. 74-83, Feb. 2015, doi: 10.1109/THMS.2014.2358634.

[19] R. N. Khushaba, E. Scheme, A. H. Al-Timemy, A. Phinyomark, A. Al- Taee, A. Al-Jumaily, "A long short-term recurrent spatial-temporal fusion for myoelectric pattern recognition", in Expert Systems with Applications, vol. 178, 2021, https://doi.org/10.1016/j.eswa.2021.114977.

[20] A. Olsson, N. Maleševic, A. Björkman and C. Antfolk, "Exploiting the Intertemporal Structure of the Upper-Limb sEMG: Comparisons between an LSTM Network and Cross-Sectional Myoelectric Pattern Recognition Methods", 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany, 2019, pp. 6611-6615, doi: 10.1109/EMBC.2019.8856648.

[21] C. Wang, W. Guo, H. Zhang, L. Guo, C. Huang, and C. Lin, "sEMG-based continuous estimation of grasp movements by long-short term memory network", *Biomedical Signal Processing and Control*, vol. 59, 2020, https://doi.org/10.1016/j.bspc.2019.101774.

[22] U. Côté Allard, E. Campbell, A. Phinyomark, F. Laviolette, B. Gosselin, , and E. Scheme, "Interpreting Deep Learning Features for Myoelectric Control: A Comparison With Handcrafted Features". *Frontiers in Bioengineering and Biotechnology*. vol. 8, 2020, doi: 10.3389/fbioe.2020.00158.

[23] W. Yang, D. Yang, Y. Liu and H. Liu, "Decoding Simultaneous Multi-DOF Wrist Movements From Raw EMG Signals Using a Convolutional Neural Network," in IEEE Transactions on Human-Machine Systems, vol. 49, no. 5, pp. 411-420, Oct. 2019, doi: 10.1109/THMS.2019.2925191.

[24] C. Lea, M. D. Flynn, R. Vidal, A. Reiter, and G. D. Hager. "Temporal Convolutional Networks for Action Segmentation and Detection," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 1003-1012, doi: 10.1109/CVPR.2017.113.

[25] O. W. Samuel et al., "Spatio-Temporal Based Descriptor for Limb

Movement-Intent Characterization in EMG-Pattern Recognition System", 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany, 2019, pp. 2637-2640, doi: 10.1109/EMBC.2019.8857638.

[26] P. Xia, J. Hu, and Y. Peng, "EMG-Based Estimation of Limb Movement Using Deep Learning With Recurrent Convolutional Neural Networks", *Artificial Organs*, vol. 42, no. 5, pp. E67-E77, 2018, doi:10.1111/aor.13004.

[27] Y. Hu, Y. Wong, W. Wei, Y. Du, M. Kankanhalli, and W. Geng, "A novel attention-based hybrid CNN-RNN architecture for sEMG-based gesture recognition". *PLOS ONE*, vol. 13, no. 10:e0206049, 2018, doi: 10.1371/journal.pone.0206049.

[28] Y. Wu, B. Zheng and Y. Zhao, "Dynamic Gesture Recognition Based on LSTM-CNN", Chinese Automation Congress (CAC), Xi'an, China, 2018, pp. 2446-2450, doi: 10.1109/CAC.2018.8623035.

[29] U. Côté-Allard, G. Gagnon-Turcotte, F. Laviolette, and B. Gosselin, "A Low-Cost, Wireless, 3-D-Printed Custom Armband for sEMG Hand Gesture Recognition". *Sensors (Basel)*vol. 19, no. 12, 2019, doi:10.3390/s19122811.

[30] U. Côté-Allard et al., "Deep Learning for Electromyographic Hand Gesture Signal Classification Using Transfer Learning", in *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 4, pp. 760-771, April 2019, doi: 10.1109/TNSRE.2019.2896269.

[31] D. Josephs, C. Drake, A. Heroy, and J. Santerre, "sEMG Gesture Recognition with a Simple Model of Attention", Proceedings of the Machine Learning for Health NeurIPS Workshop, PMLR, vol. 136, pp. 126-138, 2020.

[32] S Dhar, J Guo, J Liu, S Tripathi, U Kurup, M Shah, "On-device machine learning: An algorithms and learning theory perspective", arXiv preprint arXiv:1911.00623.

[33] A. A. Adewuyi, L. J. Hargrove, T. Kuiken, "Evaluating EMG Feature and Classifier Selection for Application to Partial-Hand Prosthesis Control ", Frontiers in Neurorobotics, vo. 10, pp. 15, 2016. https://doi.org/10.3389/fnbot.2016.00015.

[34] O. W. Samuel et al., "Characterizing Multiple Patterns of Motor Intent Using Spatial-Temporal Information for Intuitively Active Motor Training in Stroke Survivors," 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), 2020, pp. 3831-3834, doi: 10.1109/EMBC44109.2020.9176308.

[35] B. Hjorth, "EEG analysis based on time domain properties," Electroencephalography and Clinical Neurophysiology, vol. 29, no. 3, pp. 306-310, 1970.

[36] S. Pizzolato, L. Tagliapietra, M. Cognolato, M. Reggiani, H. Müller, and M. Atzori, "Comparison of six electromyography acquisition setups on hand movement classification tasks". *PLOS ONE*. vol. 12. no. e0186132, 2017. doi: 10.1371/journal.pone.0186132.

[37] A. Krasoulis, I. Kyranou, M. Erden, K. Nazarpour, and S. Vijayakumar, "Improved prosthetic hand control with concurrent use of myoelectric and inertial measurements". *Journal of NeuroEngineering and Rehabilitation*, vol. 14, no. 71, 2017. https://doi.org/10.1186/s12984-017-0284-4

[38] Y. Geng et al., "A Robust Sparse Representation Based Pattern Recognition Approach for Myoelectric Control," in IEEE Access, vol. 6, pp. 38326-38335, 2018, doi: 10.1109/ACCESS.2018.2851282.

[39] 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, Jan. 1993, doi: 10.1109/10.204774.

[40] D. Cai, X. He and J. Han, "SRDA: An Efficient Algorithm for Large-Scale Discriminant Analysis," in IEEE Transactions on Knowledge and Data Engineering, vol. 20, no. 1, pp. 1-12, Jan. 2008, doi: 10.1109/TKDE.2007.190669.

[41] R. Menon, G. Di Caterina, H. Lakany, L. Petropoulakis, B. A. Conway and J. J. Soraghan, "Study on Interaction Between Temporal and Spatial Information in Classification of EMG Signals for Myoelectric Prostheses," in IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 25, no. 10, pp. 1832-1842, Oct. 2017, doi: 10.1109/TNSRE.2017.2687761.

[42] S. Amsüss, L. P. Paredes, N. Rudigkeit, B. Graimann, M. J. Herrmann and D. Farina, "Long term stability of surface EMG pattern classification for prosthetic control," 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2013, pp. 3622-3625, doi: 10.1109/EMBC.2013.6610327.

[43] E. Scheme and K. Englehart, "Electromyogram pattern recognition for control of powered upper-limb prostheses: state of the art and challenges for clinical use". Journal of Rehabilitation Research and Development, vol. 48, no. 6, pp. 643-59, 2011. doi: 10.1682/jrrd.2010.09.0177. PMID: 21938652.

[44] C. Wang, W. Guo, H. Zhang, L. Guo, C. Huang, and C. Lin, "sEMG-based continuous estimation of grasp movements by long-short term memory network". Biomedical Signal Processing and Control, vol. 59, 101774, 2020. https://doi.org/10.1016/j.bspc.2019.101774.

[45] J. L. Nawfel, K. B. Englehart and E. J. Scheme, "A Multi-Variate Approach to Predicting Myoelectric Control Usability," in IEEE Transactions on Neural Systems and Rehabilitation Engineering, doi: 10.1109/TNSRE.2021.3094324.

Myoelectric control with fixed convolution-based time-domain feature extraction: exploring the spatio–temporal interaction

Khushaba, Rami N.; Al-Timemy, Ali H.; Samuel, Oluwarotimi Williams; Scheme, Erik J.

2022-02-24
IEEE
https://unbscholar.lib.unb.ca/handle/1882/37615