Real-Time Forecasting of sEMG Features for Trunk Muscle Fatigue using Machine Learning

Ahmad Moniri, Dan Terracina, Jesus Rodriguez-Manzano,
Paul H. Strutton, and Pantelis Georgiou

Abstract

Objective: Several features of the surface electromyography (sEMG) signal are related to muscle activity and fatigue. However, the time-evolution of these features are non-stationary and vary between subjects. The aim of this study is to investigate the use of adaptive algorithms to forecast sMEG feature of the trunk muscles. Methods: Shallow models and a deep convolutional neural network (CNN) were used to simultaneously *learn* and *forecast* 5 common sEMG features in real-time to provide tailored predictions. This was investigated for: up to a 25 second horizon; for 14 different muscles in the trunk; across 13 healthy subjects; while they were performing various exercises. Results: The CNN was able to forecast 25 seconds ahead of time, with 6.88% mean absolute percentage error and 3.72% standard deviation of absolute percentage error, across all the features. Moreover, the CNN outperforms the best shallow model in terms of a figure of merit combining accuracy and precision by at least 30% for all the 5 features. Conclusion: Even though the sEMG features are non-stationary and vary between subjects, adaptive learning and forecasting, especially using CNNs, can provide accurate and precise forecasts across a range of physical activities. Significance: The proposed models provide the groundwork for a wearable device which can forecast muscle fatigue in the trunk, so as to potentially prevent low back pain. Additionally, the explicit real-time forecasting of sEMG features provides a general model which can be applied to many applications of muscle activity monitoring, which helps practitioners and physiotherapists improve therapy.

1 Introduction

Low Back Pain (LBP) is one of the leading causes of disability around the world, affecting over 500 million people at any one time [1–3]. There is evidence that control of the trunk muscles is altered in LBP [4–10]. In particular, changes in various muscle fatigue-related features of the electrical activity produced by skeletal muscles has been shown to be a significant factor in developing LBP [11, 12]. However, the time-course and development varies between subjects, making the identification of a global characterization of muscle fatigue very difficult [13–15].

Surface Electromyography (sEMG) is a well-established electrophysiological technique that enables the extraction of characteristics such as muscle activity which can be used to investigate fatigue. These sensors are cheap, portable and non-invasive, thus, enabling a wide range of "wearable" applications [16–18]. However, they are inherently noisy and their quality is affected by various sources [19–24]. Therefore, it is necessary to extract features to reliably characterize muscle activity [25].

Typically, these features are used as the input to an 'offline' machine learning (ML) model to find the association between the features and a given outcome, such as: muscle contraction force [26], hand motion detection [27, 28] or forehead bio-signals detection (i.e. electromyography, electrooculography and electroencephalography) [29]. However, these applications do not require the model to adapt over time in order to reflect changes in the association due to physiological variations. Given this assumption, prediction accuracy may be compromised if the data sets are not large enough to generalize across time and subjects. In order to relax this assumption and pro-

*Ahmad Moniri and Dan Terracina contributed equally to this work.

*Corresponding authors: Ahmad Moniri and Dan Terracina.

*Ahmad Moniri, Dan Terracina, Jesus Rodriguez-Manzano and Pantelis Georgiou are with the Department of Electrical and Electronic Engineering, Centre for Bio-Inspired Technology, Imperial College London, SW7 2AZ, UK. Dan Terracina and Paul H. Strutton are with The Nick Davey Laboratory, Department of Surgery and Cancer, Faculty of Medicine, Sir Michael Uren Hub, Imperial College London, White City Campus, London, W12 0BZ, UK. Jesus Rodriguez-Manzano is also with NIHR Health Protection Research Unit in Healthcare Associated Infections and Antimicrobial Resistance, Department of Infectious Disease, Faculty of Medicine, Imperial College London, UK. (e-mail: ahmad.moniri13@imperial.ac.uk; daniemanuel.terracina-barcas16@imperial.ac.uk)

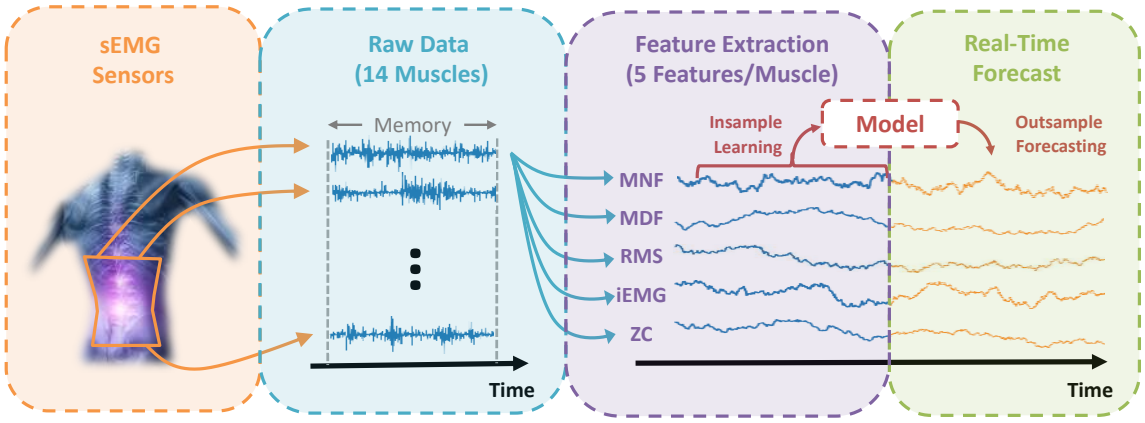


Figure 1: General workflow for forecasting features of the sEMG. First, sEMG sensors are placed on the subject which continuously record the sEMG for 14 different muscles in the lower back. These are bilaterally placed on the: latissimus dorsi, longissimus, illiocostalis, multifidus, rectus abdominis, internal and external obliques. The raw data is extracted and stored in a first-in first-out memory. Second, five features of the sEMG, namely the mean frequency (MNF), median frequency (MDF), root mean square (RMS), integrated EMG (iEMG) and zero crossings (ZC), are extracted from historical data. At every time instance, the model updates its parameters based on historical data (i.e. insample learning or training) and also uses the latest data to forecast future values of the features (outsample forecasting or testing).

vide a model which can adapt to individual subjects over time, 'online' methods are necessary.

There are many algorithms for data-driven real-time forecasting of time-series using machine learning which fall under the field of 'online learning', also referred to as 'incremental learning' or adaptive signal processing (ASP) [30]. The idea behind ASP is to update the parameters of the model through time as data is streamed such that the model naturally adapts to the latest observations. This class of models can be split into two categories: shallow learning and deep learning. The shallow models boast a low computational complexity which is ideal for wearable applications. However, their lack of computational complexity results in a lack of expressive power (i.e. inability to capture complex non-linear patterns between the inputs). This issue can be circumvented by using deep learning [31]. For example, convolutional neural networks (CNN), which are based on a stack of causal dilated convolutional layers, as in the WaveNet, have been shown to provide state-of-the-art performance in a range of tasks such as glucose time-series prediction and audio generation [32,33]. Traditionally, deep learning is not considered as a low-power solution, however, coupled with recent advances in both quantized/compressed neural networks [34,35] and AI hardware [36,37], deep solutions are expected to overcome significant challenges associated with wearable devices.

Real-time forecasting is particularly relevant for applications such as identifying the onset of muscle fatigue (between *no fatigue* and *fatigue*), where the user needs to be warned before the fatigue state, so as to potentially prevent injury. Some work has been undertaken to implicitly forecast fatigue in the biceps muscle by introducing a new state labelled as *transition to fatigue* [25,38,39]. Unfortunately, since the features are not explicitly modelled, other application areas which use the same (or similar) features cannot exploit this approach. This motivates the need for a general approach, which can be applied to many sEMG applications, that directly model and forecast the time-evolution of the well-known sEMG features in real-time.

In this work, we demonstrate accurate and precise real-time *learning* and *forecasting* for common features of the sMEG signal from a number of trunk muscles; such that it can be used for predicting muscle fatigue and preventing low back pain. In particular, this is achieved by using 5 sEMG features from 14 muscles in the trunk for 13 subjects, as depicted in Figure 1. We investigate both shallow and deep models using adaptive algorithms, which adapt to each individual subject over time; providing 'tailored' forecasts. Moreover, forecasts are investigated up to 25 seconds ahead of time which could be relevant for a wide range of applications, such as preventing injury in high intensity sports, human motor control,

physiotherapy, and exoskeletons. Once these algorithms are coupled with application specific models, and embedded into a wearable device, practitioners and physiotherapists will have access to a new tool, potentially improving therapy for subjects with LBP.

2 Data Collection

2.1 Participant Details and Tasks

Thirteen healthy male subjects participated in this study (ethics granted, ICREC ref: 19IC4970). The participants were (mean $\pm$ std): age 28.45 ± 7.92 years, height 181.1 ± 6.52 cm and weight 78.4 ± 12.3 kg. Subjects undertook a specific sequence of exercises lasting a total of approximately 70 minutes. The exercises were conducted as follows: static standing position, Maximum Voluntary Contraction (MVC) of the back & abdominal muscles, a modified Biering-Sorensen (BS) test [40], a 30 min walk, MVC of the back & abdominal muscles, the modified BS test and a plank (i.e. an isometric core strength exercise). For the MVC, participants performed 3 brief (3 s) MVC of the trunk muscles, with consistent verbal encouragement during every MVC. For the back muscles, subjects lay supine on a couch, with Velcro straps around the shoulders, hips and ankles and extended the trunk as hard as possible against the straps. For the abdominal muscles, subjects lay with the hips and knees flexed in a sit-up position. For the BS test, it involved subjects lying prone on a couch with their torso hanging off the end of the bench (with straps around the hips and ankles) and producing an isometric contraction to keep the torso parallel to the floor during 30 s [40].

2.2 Sensor Locations

Wireless sEMG sensors (Myon 320) were placed bilaterally on 7 muscles (total of 14 muscles): latissimus dorsi, longissimus, illiocostalis, multifidus, rectus abdominis, internal and external obliques, as shown in Fig. 2. Before placing the electrodes, the skin was cleaned with alcohol to get rid of skin impurities. Pairs of sEMG pre-gelled Ag-AgCl electrodes were placed according to SENIAM indications [41]. For multifidus, they were located 1 cm medially to the line joining the posterior superior iliac spine (PSIS) and the L1-L2 intervertebral space [42]. For illiocostalis, electrodes were placed 1 cm lateral to a line between the PSIS and the 12^{th} rib at the L2 vertebral level [42]. For longissimus, they were located at the L1 level, whereas latissimus dorsi was located at the T12 level along a line joining posterior axillary fold and the S2 spinous process [43, 44]. For rectus abdominis, they were placed 2 cm laterally from the umbilicus. For internal oblique, they were placed 1 cm inferiorly and medially to the anterior superior illiac spine [45]. Finally, for external oblique, they were positioned halfway between the inferior surface of the 12^{th} rib and the illiac crest at the level of the umbilicus. All pairs of electrodes were placed in the direction of the muscle fibres.

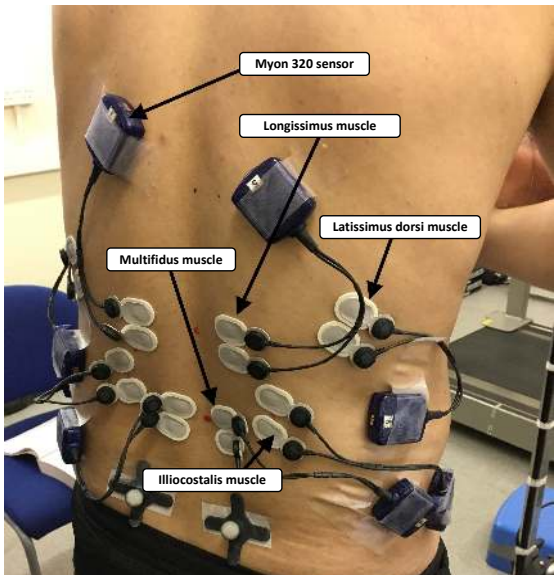


Figure 2: Position of the surface EMG sensors for the trunk muscles. Note: only the sensors placed on the back muscles (latissimus dorsi, longissimus and illiocostalis, multifidus) are displayed. Similar sensors are placed on the abdominal muscles (rectus abdominis, internal and external obliques).

3 Data Analysis

In this work, the general workflow for the real-time forecasting of muscle fatigue-related features is as follows: (i) extract raw sEMG signals from all 14 muscles; (ii) extract 5 muscle fatigue-related features from each muscle (i.e. 70 features in total); (iii) use an online algorithm to update its parameters based on the most recent observations (i.e. insample learning); (iv) and forecast all features x-steps ahead (i.e. outsample forecasting) [46].

The raw sEMG signal obtained from each sensor was sampled at 1kHz and followed the same feature extraction protocol. The features are computed us-

ing a rolling window with $N = 500$ samples. Before feature extraction, the raw sEMG signal was filtered between 20 and 450 Hz by applying a 4^{th} order Butterworth band-pass filter [47].

3.1 Feature Extraction

There are various features of the sEMG signal reported in the literature which are related to muscle activity [48]. These features can be divided into 2 categories: time-domain and frequency-domain. The time-domain features used in this work include the Root Mean Square (RMS) amplitude, integrated EMG (iEMG) and Zero-Crossings (ZC). The frequency-domain features include mean frequency (MNF) and median frequency (MDF).

3.1.1 Time-domain Features

The definition of the time-domain features, RMS amplitude, iEMG and ZC are shown in equation (1), (2) and (3) respectively. The RMS amplitude and the iEMG are related to muscle contraction strength and the motor-unit recruitment [25, 49]. From an implementation perspective, the ZC is a very attractive feature since it is trivial to implement using a simple threshold detector. Note: although the ZC is computed in the time-domain, it carries frequency information [50].

$$\text{RMS} = \sqrt{\frac{1}{N}\sum_{n=1}^{N} y_n^2} \tag{1}$$

$$\text{iEMG} = \sum_{n=1}^{N} |y_n| \tag{2}$$

$$\text{ZC} = \sum_{n=1}^{N} \left|\text{sign}(y_n) - \text{sign}(y_{n-1})\right| \tag{3}$$

Where y_n refers to the filtered EMG signal at sample n.

3.1.2 Frequency-domain Features

Regarding the frequency-domain features, the decrease in mean and median frequency of the spectrogram has been shown to be related to muscle fatigue through a decrease in pH and therefore a decrease in muscle fibre conduction velocity [51, 52].

In order to compute the features, first, the Short-Time Fourier Transform (STFT) with a Gaussian window of 500 samples is applied to the filtered sEMG signals. The MNF, as defined in (4), corresponds to the mean frequency of the power spectral density while the MDF is defined as the frequency which splits the power spectrum in half, as in (5).

$$\text{MNF} = \frac{\sum_{n=1}^{N} f_n \cdot P(f_n)}{\sum_{n=1}^{N} P(f_n)} \tag{4}$$

$$\sum_{n=1}^{\text{MDF}} P(f_n) = \sum_{n=\text{MDF}}^{N} P(f_n) = \frac{1}{2}\sum_{n=1}^{N} P(f_n) \tag{5}$$

Where f_n corresponds to the n^{th} frequency bin of the STFT and $P(f_n)$ refers to the power at f_n.

3.2 Forecasting Models

In this study, two classes of ASP models are explored based on: shallow learning and deep learning. The shallow models are variations of the Least Mean Squares (LMS) algorithm [53]. It is important to stress that this model learns in real-time. More specifically, only a single parameter update is performed at every time instant. This process is depicted in Fig 3. Moreover, the goal of this work is to evaluate the role of ASP to forecast each feature independently, therefore the joint estimation of the features and further classification is outside the scope of this paper.

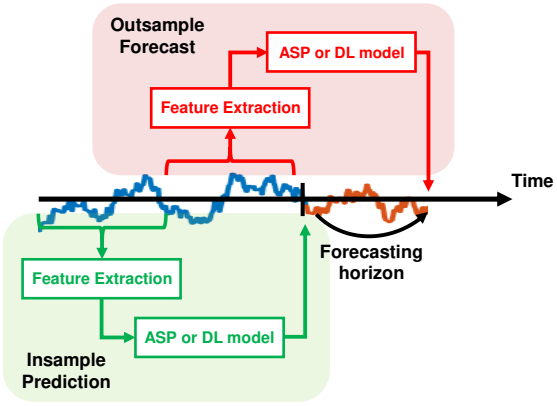


Figure 3: Forecasting procedure. At every time instance, T, a single weight update is performed based on the model learning how to predict the features at time T. Subsequently, the model, which is learning to predict X-steps ahead, uses the latest historical data in order to forecast future values.

The vanilla LMS algorithm is described in Algorithm 1. The downside of the LMS is that it is sensitive to the scaling of the input. Therefore, it is

very difficult (or perhaps impossible in certain circumstances) to choose a learning rate which guarantees stability of the algorithm [53]. To this end, many improvements to the LMS algorithm have been proposed to improve convergence and steady-state accuracy. The variations explored in this work include: Normalized LMS (NLMS), Generalized Normalized Gradient Descent (GNGD) and Affine Projection (AP) [54, 55].

Algorithm 1 Forecasting Features using LMS

Inputs:
    Feature - $\mathbf{y}$, Training window size - m, Forecasting
    horizon - l, Learning rate - μ
Output:
    Estimate of forecasted feature - $\hat{\mathbf{y}}$
Initialize $\mathbf{w} \leftarrow \text{zeros}\,(m, 1)$
while $n > m + l$
    $\mathbf{x} = \mathbf{y}[n - m - l : n - l]$
    $\hat{\mathbf{y}}[n] = \mathbf{w}^T\mathbf{x}$
    $\mathrm{e} = \mathbf{y}[n] - \hat{\mathbf{y}}[n]$
    $\mathbf{w} = \mathbf{w} + \mu \cdot e \cdot \mathbf{x}$
    $\hat{\mathbf{y}}[n + l] = \mathbf{w}^T\mathbf{y}[n - m : n]$
end while

The shallow methods mentioned above boast a very low computational complexity which is ideal for wearable or hearable applications. However, their lack of complexity means they are unable to capture complex (non-linear) patterns between the inputs.

This motivates the use of deep learning, especially since the field has seen a surge of developments in recent years; both in terms of algorithms and hardware implementation. One particular neural network architecture, which has seen state-of-the-art performance in time-series forecasting is the convolutional neural network (CNN) consisting of a stack of dilated causal convolutional layers, as depicted in Fig. 4. Therefore, the waveform, $\mathbf{y} = [y_1, y_2, \ldots, y_T]$, is directly modelled as a factorized product of conditional probabilities, as in (6).

$$\mathbf{y} = \prod_{t=1}^{T} p(y_t | y_1, y_2, \ldots, y_T) \tag{6}$$

The dilations allow the output to have a large receptive field, e.g. a network with 3 hidden layers and dilations {1, 2, 4, 8} has a receptive field of 16. The causal structure mean that the network can learn true causal relations between the inputs. The stacked convolutional architecture learns temporal patterns with better computational complexity than recurrent architectures (because of the potential for parallelization). In this study, the number of layers depends on the train window size, n_l, more specifically, the number of layers is $log_2(n_l)$. For a train window of size 64, we used a stack of 6 dilated convolutional layers with dilation factor of 1, 2, 4, 8, 16 and 32 with 2 filters of length 4 in each layer. The stack of dilated layers was followed by a fully connected layer to form a prediction. The learning algorithm used is the Adaptive Moment Estimation (ADAM) [56] optimizer and the convolutional layers adopted "He" initialization [57].

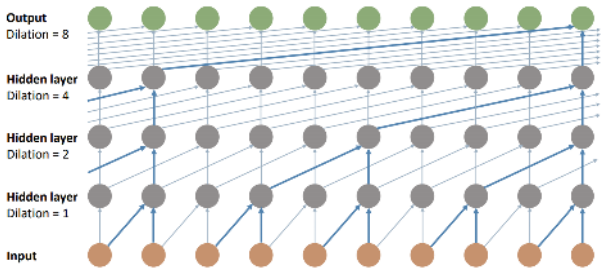


Figure 4: Visualization of a stack of dilated causal convolutional layers [33].

3.3 Performance Evaluation

The performance of the methods is evaluated using the percentage relative error, as defined in (7). This measure is applied to every data point, therefore, for every subject and feature, we can measure the accuracy by observing the mean absolute percentage error (MAPE) over time, as in (8). Similarly, we can measure the precision by computing the standard deviation of the absolute percentage error (SDAPE) over time, as in (9). Although the MAPE/SDAPE can be used to assess the relative performance between forecasting models to choose the best model, the absolute performance of a model for a given task or application requires a comprehensive study (outside the scope of this manuscript) to define an acceptable range of performance metrics.

$$E(y_n, \hat{y}_n) = 100 \times \frac{|y_n - \hat{y}_n|}{y_n} \tag{7}$$

$$\text{MAPE} = \frac{1}{N}\sum_{n=1}^{N} E(y_n, \hat{y}_n) \tag{8}$$

$$\text{SDAPE} = \sqrt{\frac{1}{N-1}\sum_{n=1}^{N}\left|E(y_n, \hat{y}_n) - \frac{1}{N}\sum_{m=1}^{N} E(y_m, \hat{y}_m)\right|^2} \tag{9}$$

In order to quantify the performance of the methods relative to each other, we can define a figure of

merit which combines the accuracy and precision into a single number. Ideally, the metric needs to capture both, the average accuracy/precision, but also the variation of accuracy and precision across subjects and muscles (since we would like a model that performs well for every feature, muscle and subject). Therefore, the figure of merit in this study is defined as the 4-product between the mean & standard deviation of the MAPE & SDAPE. Consequently, the figure of merit can be used to heuristically compare the overall performance of the models. Note: a smaller figure of merit, corresponds to a better model.

4 Results

After extracting the features for all 13 subjects, the mean length of each signal for evaluation was 8334 samples (SD: 533), with a maximum sample length of 9136 and a minimum sample length of 7593, yielding a total number of 108342 samples. For illustrative purposes, Fig. 5 (upper panel) shows an example of forecasting the MNF feature for a given subject 25 seconds ahead of time (with the LMS algorithm). Note: the outsample forecast has been shifted by 25 seconds in order to align with the value it was attempting to forecast. As expected, the insample prediction is closer to the true value than the outsample forecast. Furthermore, since the algorithm learns in real-time, there is a transient period, such that the algorithm converges to a meaningful forecast. Fig. 5 (lower panel) shows the absolute percentage error over time and depicts the measure of accuracy (MAPE) and precision (SDAPE) used in this study.

It is important to understand that each subject-muscle-feature has a value for MAPE and SDAPE. Since the distribution of errors for the 5 features are different, and it is the goal of this work to forecast individual features, it is not appropriate to aggregate their statistics. Therefore, the following results aggregate across the subjects and muscles in order to assess the performance of the models for different subjects and muscles.

Fig. 6 shows the distribution of the MAPE for all the models and features. We can clearly observe that all of the models converged to reasonable values. However, as expected, this was not the case for the naive LMS in certain circumstances, given the constant learning rate and sensitivity to scaling of the input. The mean and standard deviation of the MAPE, which is a measure of accuracy, is summarized in Table 1. In terms of average MAPE (and standard deviation of MAPE), the CNN outperformed the next best model by 7.1% (32.0%), -

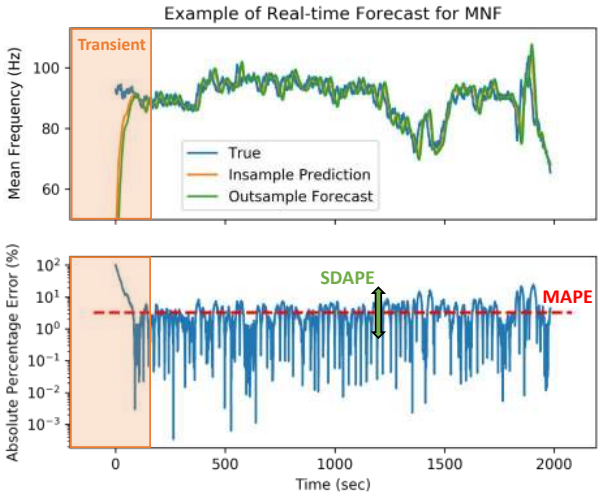


Figure 5: Example of forecasting mean frequency 25 seconds ahead using the LMS algorithm. (Upper panel) The time-series for the true realization, insample prediction and the outsample forecast. The shaded region indicates the transient period. (Lower panel) The absolute percentage error (APE) over time, along with illustrations of the MAPE and SDAPE.

3.3% (13.3%), 6.0% (3.0%), 9.0% (1.4%) and 11.1% (20.0%) for MNF, MDF, RMS, iEMG and ZC respectively. The statistical significance was confirmed using a Wilcoxon signed-rank test with Bonferroni correction ($p < 0.01$).

Table 1: Accuracy - Measured through the MAPE.

	Models				
Feature	**LMS**	**NLMS**	**GNGD**	**AP**	**CNN**
MNF	5.3 (2.5)	6.1 (5.9)	7.0 (10.3)	4.2 (2.6)	3.9 (1.7)
MDF	7.5 (3.0)	7.2 (4.0)	7.9 (7.3)	6.0 (3.0)	6.2 (2.6)
RMS	32.2 (22.8)	12.3 (7.0)	11.7 (7.1)	11.7 (6.7)	11.0 (6.5)
iEMG	31.9 (23.5)	12.2 (9.0)	11.1 (7.1)	11.2 (7.0)	10.1 (6.9)
ZC	4.4 (2.0)	5.7 (6.1)	6.2 (8.1)	3.6 (2.3)	3.2 (1.6)

Values are indicated as mean (standard deviation) over the muscles.

Table 2 shows the SDAPE for each method-feature pair. Once again, the CNN outperformed the shallow models in every case. In terms of average SDAPE (and standard deviation of SDAPE), the CNN outperformed the next best model by 14.8% (57.9%), 13.2% (28.6%), 18.2% (5.7%), 34.2% (47.1%) and 17.4% (57.1%) for MNF, MDF, RMS, iEMG and ZC respectively. The statistical significance of this result was confirmed using a Wilcoxon signed-rank test with Bonferroni correction ($p < 0.01$).

In order heuristically quantify the improvement of the deep model relative to the shallow models, Table 3 shows the figure of merit described in Section 3.3 for each of the feature-model pairs. It can be observed

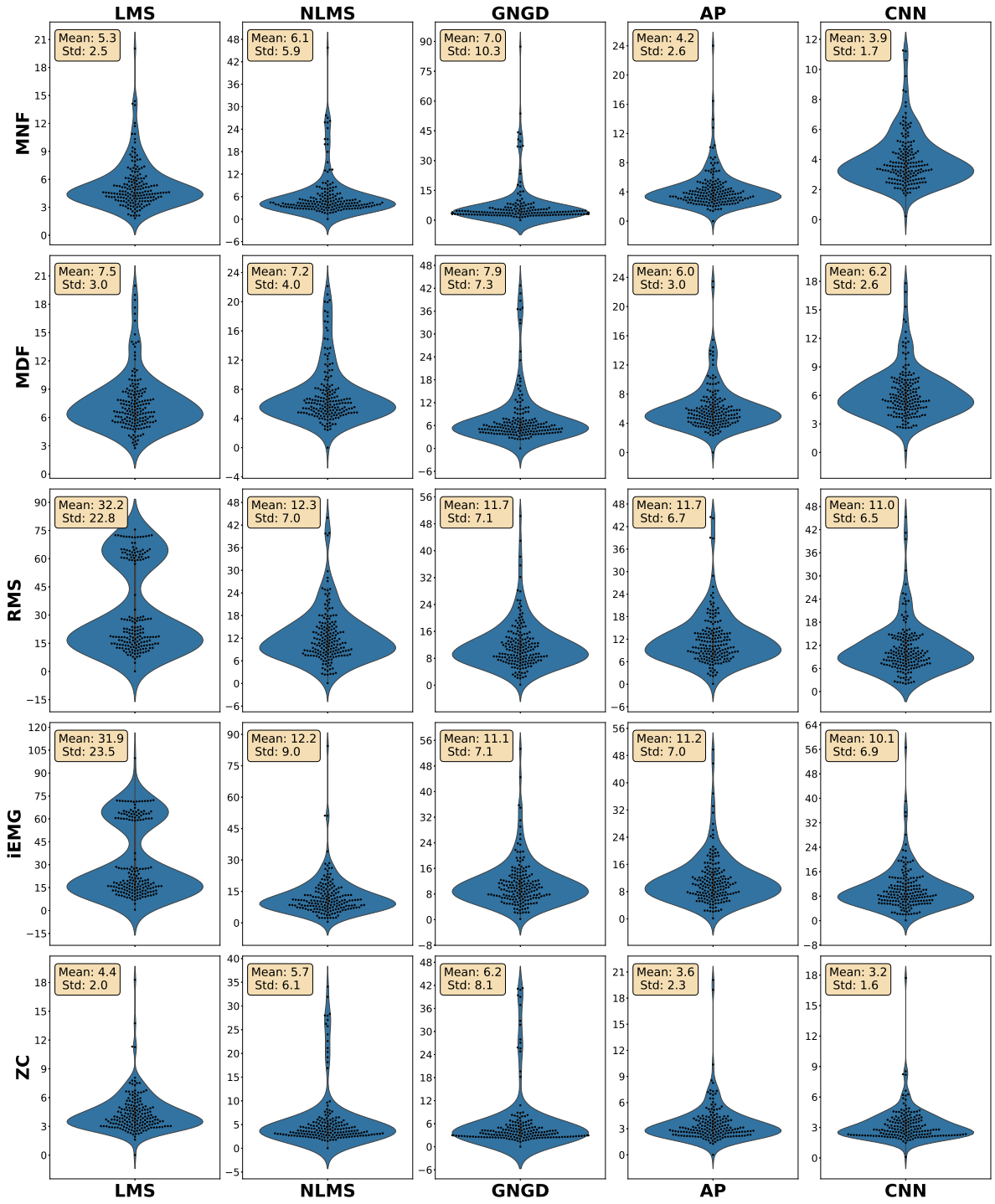
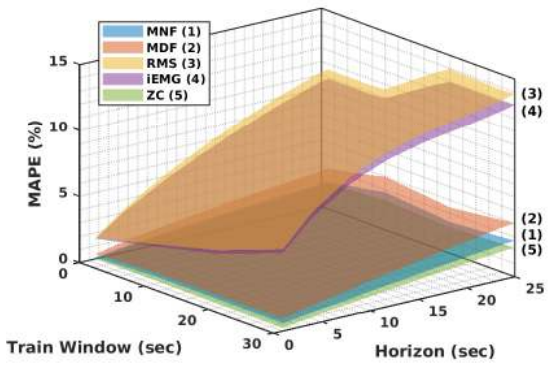
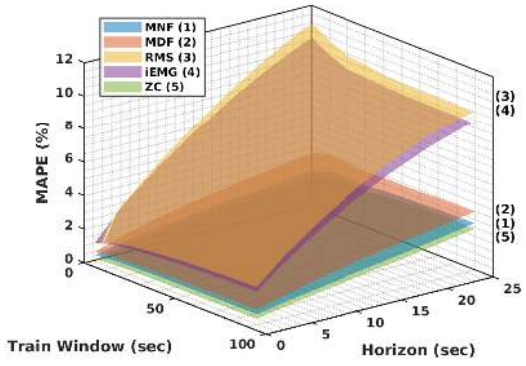


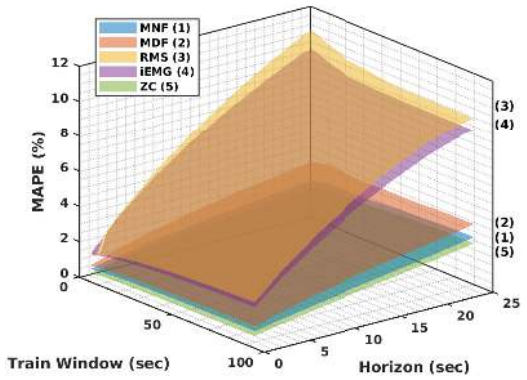
Figure 6: Violin plots showing the mean absolute percentage error (MAPE) for each model-feature pair. Every plot contains data from all 14 muscles and 13 subjects. For example, the top left plot indicates the MAPE across all muscles/patients for the mean frequency feature using the LMS algorithm. Violin plots feature a kernel density estimation of the underlying distributions. The mean and standard deviation are shown on the respective plots.



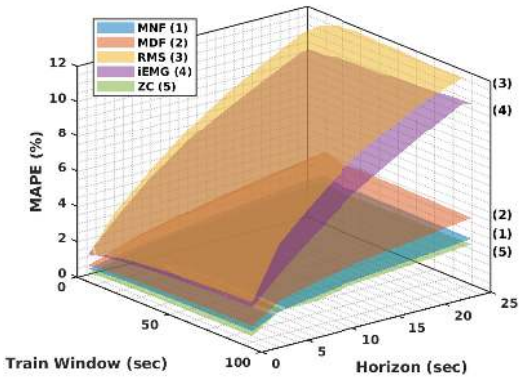
(a) LMS



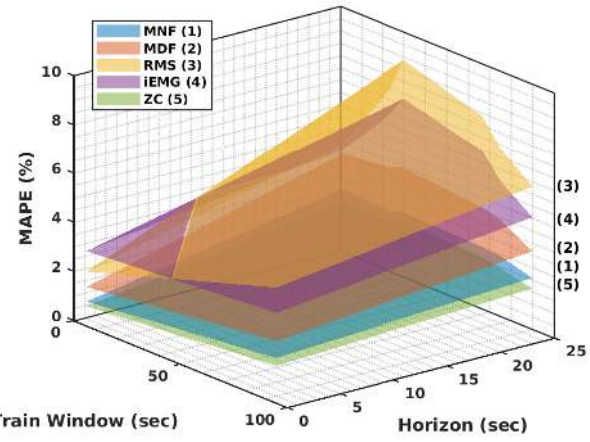
(b) NLMS



(c) GNGD



(d) AP



(e) CNN

Figure 7: Visualization of the MAPE as a function of the train window and forecasting horizon for all models and features.

Table 2: Precision - Measured through the SDAPE.

	Models				
Feature	**LMS**	**NLMS**	**GNGD**	**AP**	**CNN**
MNF	3.3 (1.9)	3.9 (3.9)	4.4 (6.6)	2.7 (2.2)	2.3 (0.8)
MDF	4.7 (2.1)	4.5 (2.8)	4.9 (4.6)	3.8 (2.4)	3.3 (1.5)
RMS	11.1 (7.8)	8.3 (5.6)	7.7 (5.5)	7.7 (5.3)	6.3 (5.0)
iEMG	10.3 (7.2)	8.1 (6.4)	7.6 (6.9)	7.3 (5.1)	4.8 (2.7)
ZC	2.7 (1.4)	3.5 (3.7)	3.8 (4.9)	2.3 (1.8)	1.9 (0.6)

Values are indicated as mean (standard deviation) over the muscles.

that the CNN improves the overall performance by 79.8%, 51.5%, 30.0%, 68.8% and 82.3% with respect to the best performing shallow model for MNF, MDF, RMS, iEMG and ZC respectively.

Table 3: Figure of Merit

	Models					
Feature	**LMS**	**NLMS**	**GNGD**	**AP**	**CNN**	**% Imp.**
MNF	83	543	2101	64	13	+79.8
MDF	222	355	1299	161	81	+51.5
RMS	63765	3982	3532	3199	1995	+30.0
iEMG	56174	5725	4125	2893	460	+68.8
ZC	34	450	926	33	4	+82.3

% Imp. = Percentage improvement of the WaveNet over the best shallow model.

The performance of the various methods is, in fact, a function of the train window size and the forecasting horizon. This relationship is shown in Fig 7 for the MAPE. Aside from the improved accuracy of the CNN, this visualization highlights 3 interesting observations: (i) For all of the models, the features which contain frequency information (i.e. MNF, MDF and ZC) had significantly smaller errors than the purely time-domain features, suggesting that the time-domain features are more difficult to forecast. (ii) The CNN reduces the disparity between the frequency-domain and time-domain features, corroborating the fact that the CNN has greater expressive power. (iii) Aside from the basic LMS, increasing the training window size improved the performance.

5 Discussion

In this work we show, for the first time, that it is possible to forecast 5 widely used features of the sEMG up to 25 seconds ahead of time across 14 different muscles in the trunk in 13 healthy subjects. We compared the performance of shallow models (LMS, NLMS, GNGD and AP) and a deep convolutional neural network. The results show that the CNN improves performance significantly in terms of both accuracy and precision. In fact, in terms of a figure of merit combining the accuracy and precision over all of the subjects and muscles, the CNN improves the overall forecasting performance by 79.8%, 51.5%, 30.0%, 68.8% and 82.8% for MNF, MDF, RMS amplitude, iEMG and ZC respectively. In order to fully understand the significance of the changes in these values during muscle fatigue, the expected change in the feature is required. However, there is currently no consensus in the literature regarding these values since it is highly related to each individual and task, and is the subject of future work [58–60].

Although the CNN has better performance over the other models, for contexts such as wearable devices, it is also important to consider the space and computational complexity. These are both intimately related to the number of parameters of the model. For a train window with 128 samples, the shallow models have exactly 128 parameters, whereas the CNN (with a single filter per layer) has 254. If we increase the number of filters per layer to 4, then we can expect the CNN to have around 10× the number of parameters than the shallow models. Fortunately, these numbers are not significant, and with recent developments in both quantized/compressed neural networks [34, 35] and AI hardware, they pose little challenge [36, 37].

In this study, time horizons up to 25 seconds were investigated, showing an average of 6.88% MAPE and 3.72% SDAPE across all features using the CNN. These results demonstrate forecasts which could be relevant for a wide range of applications. For example, as muscle fatigue develops over time, there are alterations in human motor control commands, compared to the normal (non-fatigued) state [61], resulting in a different biomechanical output. Anticipatory postural adjustments occur within tens of milliseconds and fatigue tends to affect these feed-forward mechanisms [62]. Therefore, from a clinical perspective, forecasting seconds would be appropriate to assess the subject's musculoskeletal functioning. Moreover, for many sport applications that last a few seconds or minutes and require a strenuous muscle activity, such as High Intensity Interval Training (HIIT), it is not the main interest to forecast too far in the future due to the short length of the exercise. Therefore exploring long time horizons was outside the remit of this paper.

In terms of future directions, there are many research paths that can be explored. The results presented in this paper raise a number of questions: what is the impact of the presented forecasts on subsequent classification tasks such as detecting muscle fatigue? Would the joint estimation of features improve performance? Would adding spatial information (in the form of sEMG images), in order to take advantage of spatio-temporal patterns increase the predictive performance? Can we further improve the

forecasting with different neural architectures? What is the furthest time horizon the models can accurately/precisely forecast?

In particular, increasing the dimensionality of the input via incorporating spatial information is expected to improve the performance of forecasts and enable new insights concerning muscle activity. This is because: (a) from an algorithmic perspective, there have been several recent developments in neural architectures for this type of data [63, 64]; (b) from a physiological perspective, being able to spatially localize muscle activity has significant benefits for rehabilitation and strengthening of the trunk muscles [65, 66]. For example, physiotherapists can use this information to identify weaknesses in the lower back and provide tailored therapy.

6 Conclusion

We have shown that it is possible to directly model the time evolution of the most common sEMG features related to muscle fatigue in real-time and forecast accurately and precisely up to 25 seconds ahead of time. The model with the superior performance is the deep neural network, consisting of a stack of causal dilated convolutional layers. Combined with recent advances in AI hardware, the presented models offer a low power solution for a wearable device which would be useful for practitioners, physiotherapists and the wider community for tackling the challenges of low back pain.

Acknowledgment

This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) under EPSRC DTP (EP/N509486/1 to A.M.) and the EPSRC Centre for Doctoral Training in Neurotechnology for Life and Health (to D.T).

References

[1] S. L. James, D. Abate, K. H. Abate, S. M. Abay, C. Abbafati, N. Abbasi, H. Abbastabar, F. Abd-Allah, J. Abdela, A. Abdelalim *et al.*, "Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the global burden of disease study 2017," *The Lancet*, vol. 392, no. 10159, pp. 1789–1858, 2018.

[2] D. Hoy, L. March, P. Brooks, F. Blyth, A. Woolf, C. Bain, G. Williams, E. Smith, T. Vos, J. Barendregt, C. Murray, R. Burstein, and R. Buchbinder, "The global burden of low back pain: estimates from the global burden of disease 2010 study," *Annals of the Rheumatic Diseases*, vol. 73, no. 6, pp. 968–974, 2014. [Online]. Available: https://ard.bmj.com/content/73/6/968

[3] S. Chiou, Y. Shih, L. Chou, A. McGregor, and P. Strutton, "Impaired neural drive in patients with low back pain," *European Journal of Pain*, vol. 18, no. 6, pp. 794–802, 2014.

[4] P. H. Strutton, S. Theodorou, M. Catley, A. H. McGregor, and N. J. Davey, "Corticospinal excitability in patients with chronic low back pain," *Clinical Spine Surgery*, vol. 18, no. 5, pp. 420–424, 2005.

[5] S. Chiou, A. Jeevathol, A. Odedra, and P. Strutton, "Voluntary activation of trunk extensors appears normal in young adults who have recovered from low back pain," *European Journal of Pain*, vol. 19, no. 10, pp. 1506–1515, 2015.

[6] S.-Y. Chiou, E. Koutsos, P. Georgiou, and P. H. Strutton, "Association between spectral characteristics of paraspinal muscles and functional disability in patients with low back pain: a cohort study," *BMJ open*, vol. 8, no. 2, p. e017091, 2018.

[7] P. W. Hodges, G. L. Moseley, A. Gabrielsson, and S. C. Gandevia, "Experimental muscle pain changes feedforward postural responses of the trunk muscles," *Experimental brain research*, vol. 151, no. 2, pp. 262–271, 2003.

[8] H. Tsao, L. A. Danneels, and P. W. Hodges, "Issls prize winner: smudging the motor brain in young adults with recurrent low back pain," *Spine*, vol. 36, no. 21, pp. 1721–1727, 2011.

[9] R. D'hooge, P. Hodges, H. Tsao, L. Hall, D. MacDonald, and L. Danneels, "Altered trunk muscle coordination during rapid trunk flexion in people in remission of recurrent low back pain," *Journal of Electromyography and Kinesiology*, vol. 23, no. 1, pp. 173–181, 2013.

[10] T. Nicolaisen and K. Jørgensen, "Trunk strength, back muscle endurance and low-back trouble." *Scandinavian journal of rehabilitation medicine*, vol. 17, no. 3, pp. 121–127, 1985.

[11] S. H. Roy, C. L. De, and D. A. Casavant, "Lumbar muscle fatigue and chronic lower back pain." *Spine*, vol. 14, no. 9, pp. 992–1001, 1989.

[12] S. H. Roy, C. J. De Luca, L. Snyder-Mackler, M. S. Emley, R. L. Crenshaw, and J. P. Lyons, "Fatigue, recovery, and low back pain in varsity rowers," *Med Sci Sports Exerc*, vol. 22, no. 4, pp. 463–9, 1990.

[13] B. Bigland-Ritchie and J. Woods, "Changes in muscle contractile properties and neural control during human muscular fatigue," *Muscle & Nerve: Official Journal of the American Association of Electrodiagnostic Medicine*, vol. 7, no. 9, pp. 691–699, 1984.

[14] S. C. Gandevia, R. M. Enoka, A. J. McComas, D. G. Stuart, C. K. Thomas, and P. A. Pierce, *Fatigue: Neural and muscular mechanisms*. Springer, 1995.

[15] F. A. Arnall, G. A. Koumantakis, J. A. Oldham, and R. G. Cooper, "Between-days reliability of electromyographic measures of paraspinal muscle fatigue at 40, 50 and 60% levels of maximal voluntary contractile force," *Clinical Rehabilitation*, vol. 16, no. 7, pp. 761–771, 2002.

[16] P. Georgiou and E. Koutsos, "Microelectronics for muscle fatigue monitoring through surface emg," in *CMOS Circuits for Biological Sensing and Processing*. Springer, 2018, pp. 133–162.

[17] K. Elmantawi, N. Miscourides, E. Koutsos, and P. Georgiou, "A 96-channel asic for semg fatigue monitoring with compressed sensing for data reduction," *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 1–5, 2018.

[18] E. Koutsos, V. Cretu, and P. Georgiou, "A muscle fibre conduction velocity tracking asic for local fatigue monitoring," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 10, no. 6, pp. 1119–1128, 2016.

[19] B. Afsharipour, S. Soedirdjo, and R. Merletti, "Two-dimensional surface emg: The effects of electrode size, interelectrode distance and image truncation," *Biomedical Signal Processing and Control*, vol. 49, pp. 298–307, 2019.

[20] B. Elfving, D. Liljequist, E. Mattsson, and G. Németh, "Influence of interelectrode distance and force level on the spectral parameters of surface electromyographic recordings from the lumbar muscles," *Journal of Electromyography and Kinesiology*, vol. 12, no. 4, pp. 295–304, 2002.

[21] D. Farina, R. Merletti, B. Indino, and T. Graven-Nielsen, "Surface emg crosstalk evaluated from experimental recordings and simulated signals," *Methods of information in medicine*, vol. 43, no. 01, pp. 30–35, 2004.

[22] E. Kupa, S. Roy, S. Kandarian, and C. De Luca, "Effects of muscle fiber type and size on emg median frequency and conduction velocity," *Journal of applied physiology*, vol. 79, no. 1, pp. 23–32, 1995.

[23] M. B. I. Reaz, M. Hussain, and F. Mohd-Yasin, "Techniques of emg signal analysis: detection, processing, classification and applications," *Biological procedures online*, vol. 8, no. 1, p. 11, 2006.

[24] M. Cifrek, V. Medved, S. Tonković, and S. Ostojić, "Surface emg based muscle fatigue evaluation in biomechanics," *Clinical Biomechanics*, vol. 24, no. 4, pp. 327 – 340, 2009. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0268003309000254

[25] M. R. Al-Mulla, F. Sepulveda, and M. Colley, "A review of non-invasive techniques to detect and predict localised muscle fatigue," *Sensors*, vol. 11, no. 4, pp. 3545–3594, 2011.

[26] S. P. Arjunan, D. K. Kumar, and G. Naik, "Computation and evaluation of features of surface electromyogram to identify the force of muscle contraction and muscle fatigue," *BioMed research international*, 2014.

[27] M. R. Ahsan, M. I. Ibrahimy, and O. O. Khalifa, "Emg motion pattern classification through design and optimization of neural network," in *2012 International Conference on Biomedical Engineering (ICoBE)*. IEEE, 2012, pp. 175–179.

[28] G. R. Naik, D. K. Kumar, and M. Palaniswami, "Multi run ica and surface emg based signal processing system for recognising hand gestures," in *2008 8th IEEE International Conference on Computer and Information Technology*. IEEE, 2008, pp. 700–705.

[29] L. Wei, H. Hu, and K. Yuan, "Use of forehead bio-signals for controlling an intelligent wheelchair," in *2008 IEEE International Conference on Robotics and Biomimetics*. IEEE, 2009, pp. 108–113.

[30] A. H. Sayed, *Fundamentals of adaptive filtering*. John Wiley & Sons, 2003.

[31] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, p. 436, 2015.

[32] A. v. d. Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu, "Wavenet: A generative model for raw audio," *arXiv preprint arXiv:1609.03499*, 2016.

[33] K. Li, C. Liu, T. Zhu, P. Herrero, and P. Georgiou, "Glunet: A deep learning framework for accurate glucose forecasting," *IEEE journal of biomedical and health informatics*, 2019.

[34] Y.-D. Kim, E. Park, S. Yoo, T. Choi, L. Yang, and D. Shin, "Compression of deep convolutional neural networks for fast and low power mobile applications," *arXiv preprint arXiv:1511.06530*, 2015.

[35] H. Li, S. De, Z. Xu, C. Studer, H. Samet, and T. Goldstein, "Training quantized nets: A deeper understanding," in *Advances in Neural Information Processing Systems*, 2017, pp. 5811–5821.

[36] K. Ovtcharov, O. Ruwase, J.-Y. Kim, J. Fowers, K. Strauss, and E. S. Chung, "Accelerating deep convolutional neural networks using specialized hardware," *Microsoft Research Whitepaper*, vol. 2, no. 11, pp. 1–4, 2015.

[37] M. Douthwaite, F. Garcia Redondo, P. Georgiou, and S. Das, "A time-domain current-mode MAC engine for analogue neural networks in flexible electronics," in *IEEE Biomedical Circuits and Systems Conference (BioCAS)*, 2019.

[38] M. Al-Mulla, F. Sepulveda, M. Colley, and A. Kattan, "Classification of localized muscle fatigue with genetic programming on semg during isometric contraction," in *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2009, pp. 2633–2638.

[39] M. Al-Mulla, F. Sepulveda, M. Colley, and F. Al-Mulla, "Statistical class separation using semg features towards automated muscle fatigue detection and prediction," in *2009 2nd International Congress on Image and Signal Processing*. IEEE, 2009, pp. 1–5.

[40] F. Biering-Sørensen, "Physical measurements as risk indicators for low-back trouble over a one-year period." *Spine*, vol. 9, no. 2, pp. 106–119, 1984.

[41] H. J. Hermens, B. Freriks, C. Disselhorst-Klug, and G. Rau, "Development of recommendations for SEMG sensors and sensor placement procedures," *Journal of Electromyography and Kinesiology*, vol. 10, no. 5, pp. 361–374, 2000.

[42] J. L. De Foa, W. Forrest, and H. Biedermann, "Muscle fibre direction of longissimus, iliocostalis and multifidus: landmark-derived reference lines." *Journal of anatomy*, vol. 163, p. 243, 1989.

[43] M. P. de Sèze and J. R. Cazalets, "Anatomical optimization of skin electrode placement to record electromyographic activity of erector spinae muscles," *Surgical and Radiologic Anatomy*, vol. 30, no. 2, pp. 137–143, 2008.

[44] R. De Nooij, L. A. C. Kallenberg, and H. J. Hermens, "Evaluating the effect of electrode location on surface EMG amplitude of the m. erector spinae p. longissimus dorsi," *Journal of Electromyography and Kinesiology*, vol. 19, no. 4, pp. e257–e266, 2009. [Online]. Available: http://dx.doi.org/10.1016/j.jelekin.2008.03.013

[45] E. Swinnen, J. P. Baeyens, R. Meeusen, and E. Kerckhofs, "Methodology of electromyographic analysis of the trunk muscles during walking in healthy subjects: A literature review," *Journal of Electromyography and Kinesiology*, vol. 22, no. 1, pp. 1–12, 2012. [Online]. Available: http://dx.doi.org/10.1016/j.jelekin.2011.04.005

[46] D. Terracina, A. Moniri, J. Rodriguez-Manzano, P. Strutton, and P. Georgiou, "Real-time forecasting and classification of trunk muscle fatigue using surface electromyography," in *IEEE Biomedical Circuits and Systems Conference (BioCAS)*, 2019.

[47] C. J. De Luca, L. D. Gilmore, M. Kuznetsov, and S. H. Roy, "Filtering the surface emg signal: Movement artifact and baseline noise contamination," *Journal of biomechanics*, vol. 43, no. 8, pp. 1573–1579, 2010.

[48] R. Chowdhury, M. Reaz, M. Ali, A. Bakar, K. Chellappan, and T. Chang, "Surface electromyography signal processing and classification techniques," *Sensors*, vol. 13, no. 9, pp. 12 431–12 466, 2013.

[49] R. Merletti and D. Farina, *Single-Channel Techniques for Information Extraction from the Surface EMG Signal*. Wiley-IEEE Press, 2016.

[50] N. Nazmi, M. Abdul Rahman, S.-I. Yamamoto, S. Ahmad, H. Zamzuri, and S. Mazlan, "A review of classification techniques of emg signals during isotonic and isometric contractions," *Sensors*, vol. 16, no. 8, p. 1304, 2016.

[51] L. Brody, M. T. Pollock, S. H. Roy, C. De Luca, and B. Celli, "ph-induced effects on median frequency and conduction velocity of the myoelectric signal," *Journal of Applied Physiology*, vol. 71, no. 5, pp. 1878–1885, 1991.

[52] M. Cifrek, V. Medved, S. Tonković, and S. Ostojić, "Surface emg based muscle fatigue evaluation in biomechanics," *Clinical biomechanics*, vol. 24, no. 4, pp. 327–340, 2009.

[53] S. S. Haykin, *Adaptive filter theory*. Pearson Education India, 2005.

[54] D. P. Mandic, "A generalized normalized gradient descent algorithm," *IEEE signal processing letters*, vol. 11, no. 2, pp. 115–118, 2004.

[55] A. Gonzalez, M. Ferrer, F. Albu, and M. De Diego, "Affine projection algorithms: evolution to smart and fast algorithms and applications," in *2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO)*. IEEE, 2012, pp. 1965–1969.

[56] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[57] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1026–1034.

[58] R. De Nooij, L. Kallenberg, and H. J. Hermens, "Evaluating the effect of electrode location on surface emg amplitude of the m. erector spinae p. longissimus dorsi," *Journal of electromyography and kinesiology*, vol. 19, no. 4, pp. e257–e266, 2009.

[59] R. Merletti, L. L. Conte, E. Avignone, and P. Guglielminotti, "Modeling of surface myoelectric signals. i. model implementation," *IEEE transactions on biomedical engineering*, vol. 46, no. 7, pp. 810–820, 1999.

[60] B. Elfving and A. Dedering, "Task dependency in back muscle fatigue–correlations between two test methods," *Clinical Biomechanics*, vol. 22, no. 1, pp. 28–33, 2007.

[61] F. Monjo, R. Terrier, and N. Forestier, "Muscle fatigue as an investigative tool in motor control: A review with new insights on internal models and posture–movement coordination," *Human movement science*, vol. 44, pp. 225–233, 2015.

[62] H. Tsao and P. W. Hodges, "Immediate changes in feedforward postural adjustments following voluntary motor training," *Experimental brain research*, vol. 181, no. 4, pp. 537–546, 2007.

[63] S. Xingjian, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-c. Woo, "Convolutional lstm network: A machine learning approach for precipitation nowcasting," in *Advances in neural information processing systems*, 2015, pp. 802–810.

[64] A. Ziat, E. Delasalles, L. Denoyer, and P. Gallinari, "Spatio-temporal neural networks for space-time series forecasting and relations discovery," in *2017 IEEE International Conference on Data Mining (ICDM)*. IEEE, 2017, pp. 705–714.

[65] J.-Y. Tsauo, W.-H. Chen, H.-W. Liang, and Y. Jang, "The effectiveness of a functional training programme for patients with chronic low back pain–a pilot study," *Disability and rehabilitation*, vol. 31, no. 13, pp. 1100–1106, 2009.

[66] K. L. Newcomer, T. D. Jacobson, D. A. Gabriel, D. R. Larson, R. H. Brey, and K.-N. An, "Muscle activation patterns in subjects with and without low back pain," *Archives of physical medicine and rehabilitation*, vol. 83, no. 6, pp. 816–821, 2002.