Deep Learning of EMG Time–Frequency Representations for Identifying Normal and Aggressive Actions

Haya Alaskar

Prince Sattam Bin Abdulaziz, Alkharj, KSA

Summary
Convolutional neural networks (CNN) provide an interesting model to automatically identify patterns on signals. This study presents the end-to-end deep learning derived from time-frequency representations of EMG signals to identify physical activity. End-to-end learning of CNN allows the network to automatically learn features from time-frequency representations, without requiring the design of hand-crafted expert features. This type of learning eliminates the requirement for complex multi-step machine learning processing methods. The purpose of this article is to present the framework of the end-to-end learning used to classify physical activity on EMG signals into normal and aggressive classes. This paper proposes the novel approach of using the time-frequency representations produced from EMG signals as the inputs of the CNN to identify activity patterns. The importance of selecting the optimal time-frequency analysis method to represent EMG data is investigated. Three convolutional neural networks were evaluated for two time-frequency representations: the spectrogram and the scalogram. From the analysis, it can be proven that EMG signal representation affects the performance of CNNs. Using the scalogram images to train CNNs echived higher accuracy compared to the spectrogram images. Simple CNN obtained the highest classification accuracy with 94.61%.

Key words:
Convolutional neural networks, AlexNet, pre-trained , scalogram, spectrogram.

1. Introduction

Deep learning is a branch of the artificial neural network. Deep learning has a unique hierarchical structure and the ability to extract high-level features. Deep learning networks have been used widely in a number of applications, such as medical [1], [2], industry [3], [4] and financial [5] applications. They have demonstrated their ability to deal with a variety of data types, medical image diagnoses [6], [7], biomedical signal classifications [1], [8], speech [9], [10] and fault diagnoses [3]. The extensive research and application of deep learning has motivated future research to consider deep learning as the first choice for machine learning tools. Despite their different applications, most of the studies have shown that deep learning significantly outperforms traditional machine learning [2], [6], [7]. Furthermore, the most important advantage of using deep learning is the ability to extract the optimal features of each data set automatically, without the need for an expert domain. Therefore, most of the research conducted has employed deep learning networks for end-to-end learning wherein multi-step machine learning (e.g., feature extraction and selection) will be directly implemented by deep learning. There is no need for feature engineering.

In the traditional machine learning tools, classification accuracy depends on the optimal data representation or features extracted [11], [12]. Feature engineering is the procedure of extracting, combining and employing features based on human ingenuity and expert knowledge to attain at more representative ones [1], [2], [6]. The investigation and discovery of the optimal features set for particular signals is a more complex task and time-consuming. Therefore, researchers have expended great effort in designing automatic machine learning tools that are able to overcome feature engineering issues. The deep learning network has been considered in a number of applications as a form of end-to-end learning wherein feature extraction, pre-processing and classification are conducted directly using the deep learning network. Deep learning applications automatically learn discriminant features from images. In this paper, an end-to-end learning approach will be employed to classify activity patterns from EMG signals via time-frequency images.

Electromyography (EMG) was used in this study in order to train CNNs to distinguish activity patterns. In previous studies, EMG signals have been used to record the electrical activity of muscle cells and for identifying actions [13], disease detection [14], [15], and emotion detection [16]. The literature on EMG-based activity recognition mainly expands on feature engineering, with the aim of identifying pattern EMG signals in a discriminative manner [11], [14], [15]. In this study, the EMG signals denote aggressive or normal activity.

Two linear time-frequency transformations are utilized on EMG signals as image inputs to the CNN architecture: the short-time Fourier transform (STFT) spectrogram and the wavelet transform (WT) scalogram. These types of images are another form of raw signal feature representation. The main contribution of this study is to examine the performance of different CNN architectures with time-frequency representation to classify physical activity patterns on EMG signals. Two types of time-frequency representation were selected according to [17]. They were able to provide appropriate outputs for the discovery of

Manuscript received December 5, 2018
Manuscript revised December 20, 2018

complex and high-dimensional representations [17]. They were also able to capture the unknown and hidden features embedded in signals. A number of experiments will be conducted in order to find the optimal configuration structure for deep learning. Each experiment will be evaluated by computing several performance measurements. The rest of the paper is organized as follows. The remainder of Section 2 will give a brief introduction of the related work. Section 3 presents the end-to-end learning methodology for EMG signal classification. In Section 4, the methodology is applied to four experiments and the experimental results are discussed. Section 5 addresses these discussions. Section 6 concludes the paper.

2. Related Work

Since deep learning deals efficiently with images, scientists have resorted to transforming signals into visual representations based on time-frequency representation. Time-frequency can disclose characteristic signal patterns. It is also a powerful tool for characterizing medical signals [1], [18]. Furthermore, it holds more hidden features of signals and may offer a better performance in a classification tool compared with other feature extraction methods. Examples of this approach can be found in recent research. Time and frequency representation can be generated using three types of representations associated with the Fourier transform. The Fourier transform has a giant number of variants, relying on a series of properties on the data. In the spectrogram image, the different energy values represented vary with time and different colors. Therefore, researchers use spectrogram images as CNN inputs without any feature selection or extraction procedure. In [19], spectrogram images were used to train a CNN for automatic AF detection. The CNN was trained on 8,528 ECGs and tested on 3,685 ECGs ranging from 9 to 60 seconds in length. The researchers proposed a 16-layer CNN. The classification accuracy for the CNN was 82%, meaning the proposed CNN recognized normal rhythm, AF and other rhythms with an accuracy of 90%, 82% and 75%, respectively. They concluded that a CNN is able to automatically perform ECG signal classification and further, can also possibly aid in robust patient diagnosis.

Other studies have focused on diagnosing sleep disorders, such as insomnia, narcolepsy or sleep apnea, using a CNN with time-frequency images. For example [20], a time-frequency domain was generated from EEG signals in order to classify sleep stages. Multi-taper spectral estimation was used to reduce bias and variance in spectrogram images. VGGNet was used with two approaches, first with one VGG-FE that used the network as a feature extractor. The second approach pertained to VGG-FT. The highest accuracy was achieved for VGG-FE with 89%, wherein most of the sleep stages were correctly detected, namely Slow Wave Sleep (89%) with the Rapid Eye Movement (REM) stage (81%), wake stage (78%) and N2 sensitivity (75%). However, the N1 stage was incorrectly classified at a rate of 44%.

A similar study was conducted using a CNN for automatic sleep-stage scoring based on a single-channel EEG [21]. This study used an openly available dataset on 20 healthy young adults for evaluation and applied a 20-fold cross validation. The CNN was used with stochastic gradient descent (SGD) optimization. We achieved the high mean F1-score of 81%, whereas overall accuracy regarding the sleep stages was 74%.

In addition, ECG time-frequency images have been classified using a CNN. In this study [2], the time-frequency image for a heartbeat signal was created by applying a modified frequency slice wavelet transform (MFSWT). Features were automatically extracted by the stacked denoising auto-encoder (SDA) from the time-frequency image. A DNN classifier was used to classify the heartbeat. The proposed model was evaluated based on the MIT-BIH arrhythmia database. The proposed method achieved an overall accuracy of 97.5%. The spectrogram image was produced from EEG signals in order to identify motor impairment neural disorder in a person. Each image was passed to train the CNN [22]. The CNN combined with the recurrent neural network RNN was employed to estimate kinematic information for myoelectric control from the channels of EMG signals [23]. The EMG signals were converted to the time-frequency domain as inputs to the CNN. The experimental results proved that the CNN with RNN offered higher accuracy compared with using the CNN alone.

Another study [18] attempted to analyze an EEG via spectrogram images by using a CNN. Their motivation was to identfy clinical brain death diagnosis. In this paper, a deep learning structure named “Caffe” was used to design the CNN. The EEG signals were obtained from brain-damaged patients. The EEG datasets contained 36 patients, including 19 coma patients and 17 brain-dead patients. Spectrogram images were generated from these signals using STFT. And in order to increase the number of images created, six channels of the EEG signals were used to create spectrogram images. In addition, every window of STFT overlapped by 20% with the adjacent windows.

3. Methods

In this study, three CNN architectures will be explored and evaluated and the influence of some factors, such as the learning rate and type of optimization, will be analyzed. Furthermore, two types of data representation will be evaluated: the spectrogram and scalogram images. In this study, CNN transfer learning will be used. The study experiment will be implemented using eight channels that recorded the EMG signals of 10 aggressive and 10 normal actions of three men and one woman, which were then analyzed to classify the normal and aggressive actions. Figure 1 illustrates the main approach of using end-to-end learning via two types of time-frequency representations.

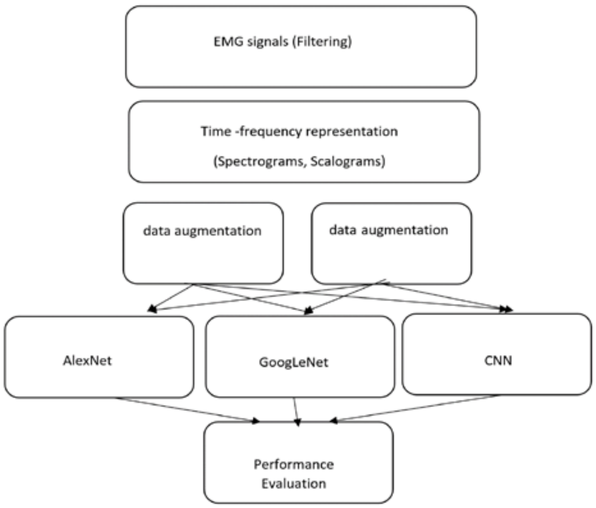

Fig. 1 The experiment procedure

3.1 EMG data set

In this paper, the EMG physical action signals from the machine learning repository (UCI) [11] was used. The signals were recorded from four subjects (3 men, 1 woman) aged 25 to 30 years. Each subject was asked to perform a different set of physical exercises. Each subject had to perform 10 normal and 10 aggressive activities. The normal activities and the aggressive activities are represented in Table 1. Eight channels were collected from eight electrodes, which corresponded to eight input time series, one for each muscle channel (ch1–8): the right bicep (ch1), right tricep (ch2), left bicep (ch3), left tricep (ch4), right thigh (ch5), right hamstring (ch6), left thigh (ch7), and left hamstring (ch8). Each time series contained about 10,000 samples, which were 10 s in length.

Table 1: The data set description

Data sets	
Subject	(3 men, 1 woman)
Electrode	8 channels
Aggressive classes	10 type of aggressive activity (elbowing, front kicking, hammering, headering, kneeing, pulling, punching, pushing, side kicking, and slapping)
Normal classes	10 types of normal activity (bowing, clapping, handshaking, hugging, jumping, running, seating, standing, walking, and waving)
Length of EMG signals	10,000 samples

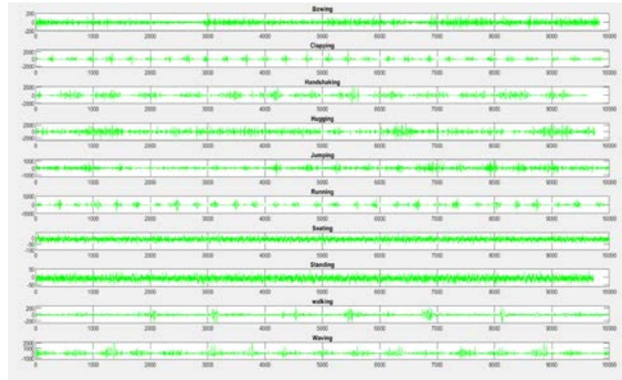

Fig. 2(a) EMG Normal Actions

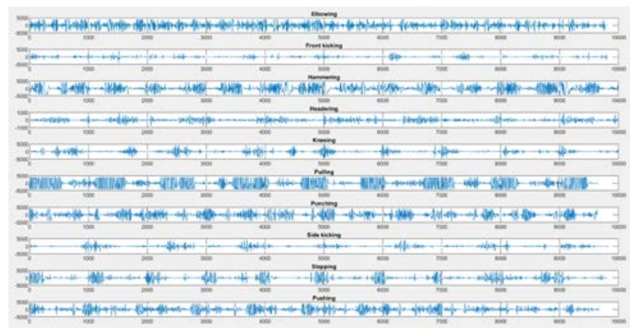

Fig. 2(b) EMG Aggressive Actions

As shown in Figure 2, the EMG signals in Figure 2(a) that presented normal actions for channel 1 is quite different than EMG signals for aggressive actions which is illustrated in Figure 2(b).

3.2 Data preparation

The experiment procedure is shown in Figure 1. First, the raw EMG signals of each subject were downloaded from [18]. The EMG data was pre-processed to prepare for generating two types of time-frequency representations, namely the spectrogram and scalogram images. Since recording signals can be affected by a number of noises, which will badly impact the signals, filtering signals were applied to reduce the noise. First, a first-order 1-Hz low-pass Butterworth filter was applied as recommended in [10]. In order to increase the number of generated images the CNN input, each signal was divided into three segments, with the 20% that overlapped each part used to produce one spectrogram image. In addition, eight channels of the EMG signals were used to create spectrogram images. The individual differences of each subject were discounted, so the images generated from different subjects were taken as one class of the dataset.

Overall, the EMG data was used to create 1,920 spectrogram or scalogram images, including 960 images generated from normal activity and 960 images generated from aggressive activity. In the next section, time-frequency representation will be highlighted and briefly introduced.

3.2.1 Time-frequency representation

Time-frequency characterizes a signal in both the time and frequency domains. Different types of methods can be used to extract time-frequency representations and the most popular types are spectrograms and scalograms. A spectrogram is a visual time-frequency representation of the signal using the STFT, whereas a scalogram uses the WT. The main difference between the two methods is that spectrograms have a fixed frequency resolution based on the window's size, whereas scalograms have a frequency based on frequency resolution.

In this study, the effectiveness of these two types of time-frequency representation will be evaluated. These types of time-frequency representations will be extracted as 2D images and passed into the CNN.

3.2.1.1. Spectrograms: Short-Time Fourier Transform (STFT)

Spectrograms are generated using the STFT. They can be represented as 2D images where the x-axis represents time and the y-axis represents frequency, and the color scale of the image shows the amplitude of the frequency. The STFT representation is based on a series of sinusoidal functions. The frequency spectrum represented on the spectrogram image varies with time. Different colors on the spectrogram image show different energy values.

A spectrogram image holds more unknown features of EMG signals and it can achieve a better performance in classification tools [11], [23]. In order to create spectrogram images from EMG signals, a Matlab function was created to compute the STFT.

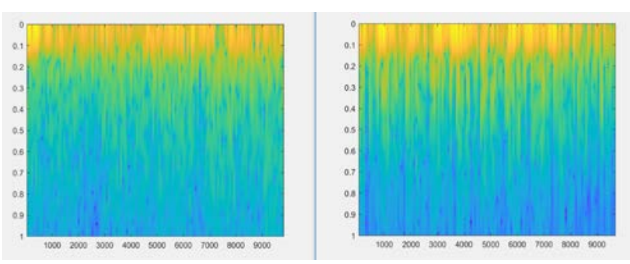

Fig. 3(a) Spectrogram image for Normal action

Fig. 3(b) Spectrogram image for Aggressive action

3.2.1.2. Scalograms: Wavelet Transform (WT)

Scalograms represent the WT. WTs are extracted based on the wavelet instead of sinusoidal functions. The WT is an optimal method for non-stationary and transient signals [25].

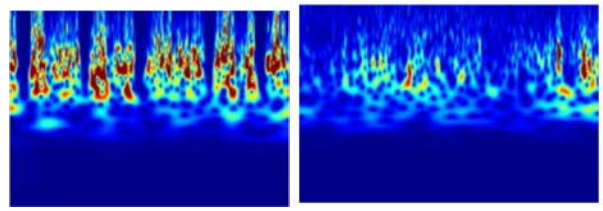

Fig. 4(a) Scalograms image for Aggressive Action

Fig. 4(b) Scalograms image for Normal Action

3.3 Convolutional Neural Network Architectures

The architecture of a convolutional neural network differs from that of a traditional artificial neural network (ANN). The CNN involve three main types of layers, namely the convolutional layer, pooling layer and fully connected layer as been illustrated in Figure 6. Each of these layers is represented as a block that contains the number of layers. Convolutions and pooling operations are employed on the input data with the use of a filter to produce an optimal feature map. In the end, these feature maps are put together as the final output of the convolution layers. Finally, classification is performed in the last fully connected layers [9].

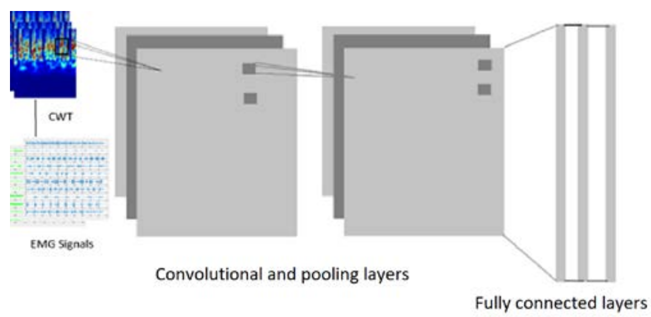

Fig. 5(a) Convlution Networks via scalograms

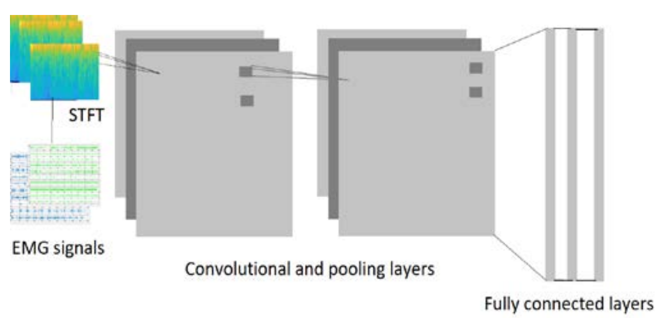

Fig. 5(b) Convolution network via spectrogram

A number of CNNs are discussed in the literature, such as AlexNet, VGGNet, GoogLeNet and ResNet. But AlexNet are commonly used for end-to-end learning and achieve very good results. For example, AlexNet have also been used to recognize cells infected with malaria and achieved 98.13% and 95.79% accuracy, respectively. Conversely, traditional machine learning tools, including support vector machine SVMs, obtained a lower accuracy of 91.66% [26]. However, these networks are pre-trained networks such as ALexNet. They were used to classify 1,000 possible categories after training on millions of images to achieve low error rates. In this study, this CNNs will not be trained from scratch; alternative pretrained CNNs with sufficient fine-tuning will be used. Furthermore, another experiment will be considered by running a simple CNN to learn from the EMG signals.

3.3.1 Convolutional layers

The convolutional layer operates on the input data by using a convolution algorithm and creating a feature map that holds the convolution calculation outputs from the previous layers. These layers performed as feature extractors to find the high-level features. So, each unit in a convolutional layer is organized in the feature maps, whereas each unit is attached to local patches in the feature maps of the previous layer with a set of weights called a filter bank. These weights will be summed up and passed to a non-linearity activation function, such as ReLU. However, the whole units associated with the feature map share the same filter bank.

3.3.2 Pooling layers

Though the task of the convolutional layer is to find the optimal features that resulted from the previous layer, the task of the pooling layer is to combine semantically similar features into select the proper feature [17]. A typical pooling unit calculates the maximum of a local area of units in feature maps. Neighboring pooling units receive input from areas that are shifted by more than one row or column, so they will reduce the number of the feature and create an invariance to small modifications and distortions. Two or three stages of convolution, non-linearity and pooling are joint, and followed by more convolutional and fully connected layers.

3.3.3 Fully connected layers

The final fully connected layer is employed to perform classification. After the data passes through multiple convolutional layers and pooling layers, the size of the output feature maps is decreased. For the classification layer, every feature map comprises only one neuron and converts to one feature vector. The feature vector is fully connected with a classifier. Usually, these layers perform as a traditional fully connected neural network.

3.4 Pretrained network

The trained CNNs were expansively trained on the large-scale, well-organized ImageNet. ImageNet is a Large-Scale Annotated Natural Image Dataset ImageNet [1] that contains more than 1.2 million images classified into 1,000 object categories. Each class has more than 1,000 images. The database is ordered according to the WordNet [55] hierarchy. Some examples of object categories in ImageNet are “sandwich,” “vase,” “cup” and so forth. ImageNet is considered the largest image dataset for visual recognition.
The aim of employing pretrained CNNs is that they can be transferred to efficiently identify EMG activity images since they were trained on large-scale images. On the other hand, learning CNN architectures (e.g., AlexNet) from scratch required tens of millions of free parameters to train, and hence a sufficiently large size of labeled images was needed, whereas recording large numbers of signals is so challenging. Therefore , applying pretrained network was recommended by a number of studies [28]–[30], wherein all CNN layers except the last are fine-tuned.
In this study, The pre-trained AlexNet architecture will be used. AlexNet was proposed in [27] and won the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in 2012. This success marked the revolution of the CNN in computer vision. The network has 11x11, 5x5 and 3x3 convolutions, max pooling, dropout, and fully connected layers. AlexNet was used to experiment with 25 layers.

3.5 Simple CNN

Simple CNN is network that learn from scratch all parameters are update according to training set. It contains 3 convolutionl layers , 3 pooling layers and 2 fully connected layer. Each output of convolution layers will be normalized by learning the data mean and variance and bass it to the next layers.

3.6 Data augmentation

Data augmentation is usually used to overcome overfitting on CNN and imitate the size data set, whereby the amount of training data is increased by using information from the original training data. The field of data augmentation has been addressed in a number of works. Data augmentation was developed by Tanner and Wong to improve simulation and make it more reasonable and simple [28], [31]. In the CNN build-up, using the number of parameter data augmentations plays a vital role in generating enough data to attain a satisfactory performance [28]. Previous research has confirmed the effectiveness of data augmentation by using simple techniques, such as rotating and flipping images [28], [32]. In this study, rotating and flipping methods have been used to generate extra data sets from the original data.

4. Experiments

End-to-end learning approaches were conducted in order to automatically solve the classification problem without needing multi-step machine learning tools. This section describes the experiments carried out to evaluate the two types of pre-trained CNNs and one simple CNN to learn directly from EMG time-frequency representation. The performance of each CNN was estimated using six types of evaluation measurements: sensitivity, specificity, accuracy, and Precision. The first two quantities were calculated by finding the true positives, true negatives, false positives and false negatives. True positives were correctly detected as aggressive. True negatives were correctly detected as normal. False positives and false negatives were the number of uncorrected aggressive and normal detections, respectively. Table 2 showed the confusion matrix to compute the evaluation measurements.

Table 2: The Confusion Matrix

	Predicted Positive	Predicted Negative
Actual Positive	TP (True Positive)	FN (False Negative)
Actual Negative	FN (False Positive)	TN (True Negative)

based on the confusion matrix in Table 2, the five measurements can be computed as follows:

$$Sensitivity = TP/(TP+FN) \quad (1)$$
$$Specificity = TN/(TN+FP) \quad (2)$$
$$Accuracy = (TP+TN)/(TP+TN+FP+FN) \quad (3)$$
$$F1=2TP/(2TP+FP+FN) \quad (4)$$
$$Precision = TP/(TP+FP) \quad (5)$$

The training options were changed in order to fit the problem. The size of the mini-batch, which is a subset of the training set to be used in each iteration of the experiment, was set at 20. Max epochs, which represent the maximum number of epochs to be used in training, were also set at 10. In this study, two learning rates were used [0.001, 0.0001] in order to estimate the best learning rate. The root mean square propagation was used as an optimizer for both CNNs. This experiment used MATLAB 2018. For the purposes of reproducibility, the network was trained in a standalone system with an Intel Core Processor i7-7500U CPU that had 2.70 GHz, 2904 MHz, two cores and 64 GB of RAM.

4.1 Result

In this study, four training experiments were performed in order to classify the patterns on EMG signals as either normal or aggressive class. Each experiment consisted of training pre-trained AlexNet, and a simple CNN.
In the initial step, each signal was filtered to removing noise. Then time-frequency representation was used to extract two types of images. Each image was fed into the CNN input. For each CNN, the training-testing rate was randomly chosen as 80% to 20% as recommended by [33]. In the next sections, the results of each experiment will be presented.

4.1.1 The spectrogram images

In this section, the results of each CNN will be represented in Table 3. Each spectrogram images ware fed into the CNN as an input. From the Table 4, it can be observed that pre-trained AlexNet performance was good compared to another CNN. It achieved high accuracy with sensitivity and specificity. Furthermore, the performance increased when the learning rate decreased. Both CNNs achieved the best result with LR = 0.0001.

Table 3: The performance of simple CNN for the Spectrogram

	Lr = 0.001		Lr = 0.0001	
	Training	Testing	Training	Testing
Accuracy	61%	61%	99.35%	69.59%
F1	0.8878	0.5769	0.9916	0.7333
SV	81.58%	0.4688	0.9935	0.7448
SP	98.34%	0.8000	0.9867	0.6333
Precision	0.9842	0.7500	0.9896	0.7222

Table 4: The performance of AlexNet for the Spectrogram

	Lr = 0.001		Lr = 0.0001	
	Training	Testing	Training	Testing
Accuracy	60.00%	56.14%	86.67%	73.91
F1	0.7181	0.7191	0.8878	0.811
SV	1.00	1.00	0.8086	1
SP	0.00	0.00	0.9834	0.4066
Precision	0.5602	0.5614	0.9842	0.6822

4.1.2The scalogram images

In this section, the results of each CNN will be represented in Tables 5. Each scalogram image was passed into the CNN as an input. From the Tables 5, it can be observed that the performance result for both CNNs were improved when scalogram mages was been used to train CNNs. However, from table 6, the pre-trained AlexNet performance was good compared to simple CNN. The high sensitivity and specificity were achieved.. Furthermore, the performance increased when the learning rate decreased. It achieved the best result with LR = 0.0001.

Table 5: The Simple CNN' Performance Results in Scalogram Images

	Lr = 0.001		Lr = 0.0001	
	Training	Testing	Training	Testing
Accuracy	95.96%	77%	100%	84.17%
F1	0.9710	0.8426	1	0.8824
SV	0.9805	0.8646	1	0.8594
SP	0.9133	0.5814	1	0.8023
Precision	0.9617	0.8218	1	0.9066

Table 6 : ALexNet' Performance Results in Scalogram Images

	Lr = 0.001		Lr = 0.0001	
	Training	Testing	Training	Testing
Accuracy	69.70%	60%	100%	87%
F1	0.7648	0.7082	0.9803	0.9096
SV	0.7150	0.7013	0.9701	0.8906
SP	0.6570	0.3768	0.9798	0.8488
Precision	0.8221	0.7152	0.9907	0.9293

4.1.3 Data augmentation

To test the effectiveness of augmentation technique, we ran two experiments on EMG signals. The results of the experiments are presented in the following Tables 7-10 for both spectrogram and scolgram images , respectively. The highest test accuracy at all the epochs was reported as the best score. It can be observed from the Table 6 that the performance of the simple CNN increased. The increasing size of the data set enhanced the CNN's performance with suitable information to increase its classification ability.

Table 7: Simple CNN' Performance for spectrogram images with data augmentation

	Lr = 0.001		Lr = 0.0001	
	Training	Testing	Training	Testing
Accuracy	76.3%	63.45	99.35%	84.8%
F1	0.8221	0.7444	0.9948	0.7960
Sv	0.9870	0.9479	0.9935	0.83330
SP	0.4718	0.2333	0.9950	0.6667
Precision	0.7045	0.6128	0.9961	0.7619

Table 8: AlexNet's performance for spectrogram images with data augmentation

	Lr = 0.001		Lr = 0.0001	
	Training	Testing	Training	Testing
Accuracy	50%	43.8%	87.77%	83.09%
F1	0	nan	0.9033	0.8646
Sv	0	0	0.8290	0.7813
SP	1	1	0.9856	0.9419
Precision	nan	nan	0.9922	0.9677

Table 9: Simple CNN' performance of scelgram images with data augmentation

	Lr = 0.001		Lr = 0.0001	
	Training	Testing	Training	Testing
Accuracy	99.62%	78.8%	100%	94.61%
F1	0.9970	0.8499	0.9980	0.8939
Sv	0.9955	0.8698	0.9987	0.9219
SP	0.9974	0.6062	0.9942	0.6860
Precision	0.9985	0.8308	0.9974	0.8676

Table 10: AlexNet' performance for sclogram images with data augmentation

	Lr = 0.001		Lr = 0.0001	
	Training	Testing	Training	Testing
Accuracy	68.95%	69.92%	93%	91%
F1	0.8162	0.8167	0.9521	0.9367
Sv	1	1	0.9376	0.9154
SP	0	0	0.9289	0.9130
Precision	0.6895	0.6901	0.9670	0.9591

5. Discussion

This study presents the novel approach of using spectrogram and scalogram images produced from surface EMG signals as the input dataset of CNNs. A deep CNN was trained to classify the physical activity patterns on EMG signals into normal and aggressive categories. The simple architecture of the CNN led to the execution of several tests to evaluate the effects of data representation, the learning rate and data augmentation. From the results of the experiments represented above, it can be observed that each CNN performed differently in each experiment. In general, applying a CNN for end-to-end learning is sufficient compared with the traditional machine learning methods. The CNN was able to achieve a high performance, especially with scalogram images. Despite offering the simplest frequency domain analysis, the spectrogram cannot sufficiently model time differences and transient signals.

Furthermore, it could be observed that the augmentation approach significantly enhanced the CNN's performance, as shown in Figure 4. In general, the augmentation method can lead to genrate the mass generation, as asserted by [28]. For example in this study circular patterns were artifacts added to generate images. A significant improvement in the CNN performance could be observed. The generated images increased the ability of the simple CNN to classify images with sufficient accuracy. Moreover, the implementation of data augmentation techniques was effective in avoiding over-fitting to some degree and the gap between training and testing accuracy increased.

However, In ALexNet's performance as illustrated in Figure 5, there is no improvement compared to the simple CNN. This is might related to that simple when CNN learn from scratch with sufficient amount of data has significantly improved simple CNN ability to learn from data to detect the right class in testing dataset. In case of pre-treained ALexNet, where network does not learn from scratch, the increasing of data does not enhance the network performance. Furthermore, the issue of overfitting has been decreased in both simple CNN and ALexNet specially when learning rate =0.0001.

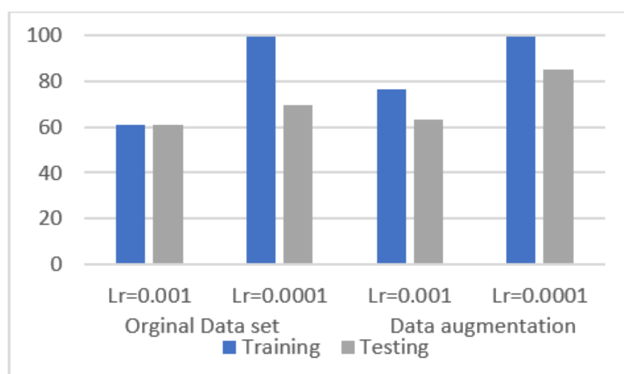

Fig. 6 Simple CNN' performance results for spectrogram images

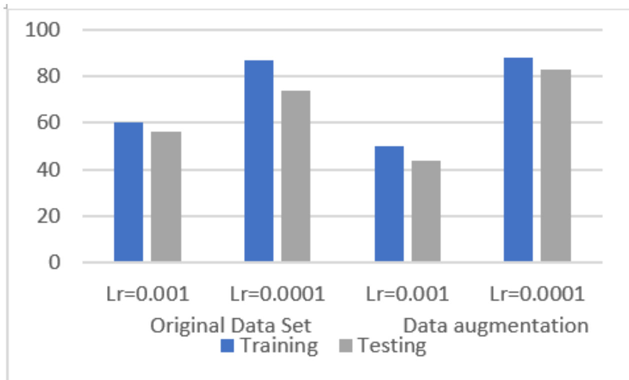

Fig. 7 AlexNet's performance results for spectrogram images

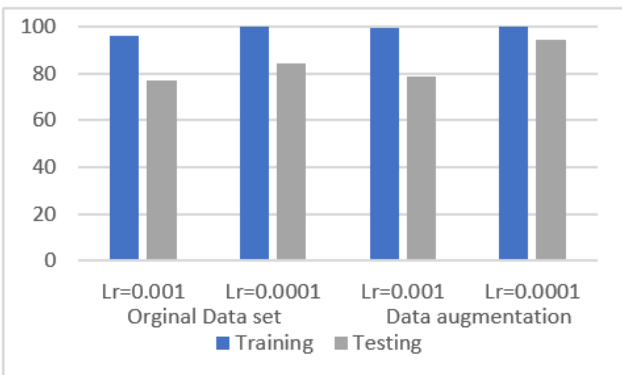

Fig. 8 simple CNN's accuracy result for sclogram images

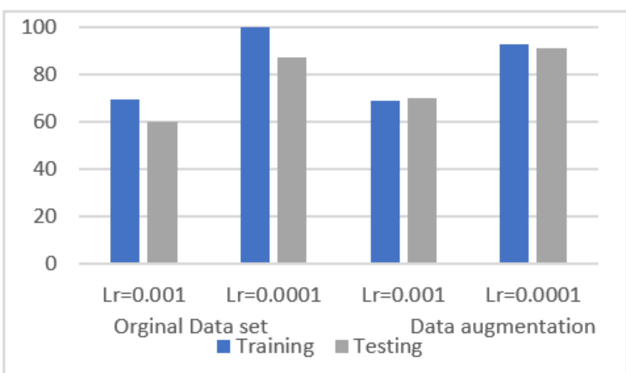

Fig. 9 AlexNet's accuracy result for sclogram images

In order to evaluate the effectiveness of transfer learning, a simple CNN was applied in this study. The experiments explored pretrained CNN architectures and a simple CNN. The results show that using a simple CNN to learn from scratch does not provide any improvements to the accuracy as shown in Figures 4. From Figure 5-7, We found that pre-trained CNNs made the gradients drop faster than those in the sample CNN in which the network was trained from scratch.

The classification accuracy was compared with a set of traditional machine learning methods used on the same datasets. For example, Amenh et al., 2017 [28] achieved 92% accuracy using random forest with wavelet packet decomposition (WPD). They used a number of feature extraction and feature reduction models in order to achieve the best performance. They applied decision tree classifiers, such as CART, random forest, C4.5 and rotation forest. Nevertheless, the results in this study were lower than the results achieved with the best reference approaches in our tests. They used the bispectrum and the quadratic phase coupling of each EMG episode was determined. However, the present study is novel, showing that CNNs have the ability to automatically extract and evaluate a set of the optimal features. This is related to the highly complex structure of CNN architectures and the fact that the present tests were designed to identify the features that help to distinguish between the two classes of EMG signals. To conclude, the viability of these two pretrained CNNs for physical activity pattern detection was fully demonstrated in terms of specificity, sensitivity, accuracy and Precision. Though, some EMG signals for aggressive actions show similar behaviors with EMG normal actions as been illustrated in Figure 2.

The requirement and the importance of monitoring human activity has been addressed in a number of research studies [34].This study can provide a suitable approach for monitoring human activity and it can be used to ensure the safety of human health, especially when people are performing their exercises to evaluate the type of exercise that they are able to do. In addition, the findings can be used to build a wearable sensor to monitor physical activity.

6. Conclusion

It can be concluded that end-to-end learning can be used to automatically analyze the EMG signals from small datasets. The results show that convolutional neural networks with a very simple architecture can create accurate results comparable to the traditional machine learning methods with feature engineering eliminated. Aggressive and normal EMG signals were analyzed using a CNN with time-frequency representation. The best performance was obtained using simple CNN with data augmentation, which offered accuracy of 94.61%. Furthermore, in this study, a comparison of two types of time-frequency representation was performed in the EMG signals. The spectrogram and scalogram images of the aggressive and normal activities of the EMG were generated each episode and fed into the inputs of the CNN. Thus, the scalogram of the EMG signal was applicable to discrete aggressive and normal activities. This effective method could help physical specialists in defining aggressive activities to monitoring patients. This can help the healthcare industry to improve activity monitoring on EMG signals. In future work, we plan to investigate the utilization of time-frequency representations (e.g., HTT) as a preprocessing step, as well as more complex CNN architectures.

References

[1] G. Ruffini et al., "Deep learning with EEG spectrograms in rapid eye movement behavior disorder," bioRxiv, p. 240267, 2018.

[2] K. Luo, J. Li, Z. Wang, and A. Cuschieri, "Patient-specific deep architectural model for ecg classification," J. Healthc. Eng., vol. 2017, 2017.

[3] S. Guo, T. Yang, W. Gao, and C. Zhang, "A Novel Fault Diagnosis Method for Rotating Machinery Based on a Convolutional Neural Network," Sensors, vol. 18, no. 5, 2018.

[4] M. He, D. He, J. Yoon, T. J. Nostrand, J. Zhu, and E. Bechhoefer, "Wind turbine planetary gearbox feature extraction and fault diagnosis using a deep-learning-based approach," Proc. Inst. Mech. Eng. Part O J. Risk Reliab., p. 1748006X18768701, 2018.

[5] W. Bao, J. Yue, and Y. Rao, "A deep learning framework for financial time series using stacked autoencoders and long-short term memory," PloS One, vol. 12, no. 7, p. e0180944, 2017.

[6] Y. Yuan and M. Q.-H. Meng, "Deep learning for polyp recognition in wireless capsule endoscopy images," Med. Phys., vol. 44, no. 4, pp. 1379–1389, 2017.

[7] D. C. Cireșan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Mitosis detection in breast cancer histology images with deep neural networks," in International Conference on Medical Image Computing and Computer-assisted Intervention, 2013, pp. 411–418.

[8] S. Chauhan and L. Vig, "Anomaly detection in ECG time signals via deep long short-term memory networks," in Data Science and Advanced Analytics (DSAA), 2015. 36678 2015. IEEE International Conference on, 2015, pp. 1–7.

[9] G. Hinton et al., "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," IEEE Signal Process. Mag., vol. 29, no. 6, pp. 82–97, 2012.

[10] S. Thomas, S. Ganapathy, G. Saon, and H. Soltau, "Analyzing convolutional neural networks for speech activity detection in mismatched acoustic conditions," in Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on, 2014, pp. 2519–2523.

[11] H. M. Alaskar, "Dynamic self-organised neural network inspired by the immune algorithm for financial time series prediction and medical data classification," PhD Thesis, Liverpool John Moores University, 2014.

[12] H. Alasker, S. Alharkan, W. Alharkan, A. Zaki, and L. S. Riza, "Detection of kidney disease using various intelligent classifiers," in Science in Information Technology (ICSITech), 2017 3rd International Conference on, 2017, pp. 681–684.

[13] A. A. Abdullah, A. Subasi, and S. M. Qaisar, "Surface EMG signal classification by using WPD and ensemble tree classifiers," in CMBEBIH 2017, Springer, 2017, pp. 475–481.

[14] G. Biagetti, P. Crippa, S. Orcioni, and C. Turchetti, "Surface EMG fatigue analysis by means of homomorphic deconvolution," in Mobile Networks for Biometric Data Analysis, Springer, 2016, pp. 173–188.

[15] G. Bovi, M. Rabuffetti, P. Mazzoleni, and M. Ferrarin, "A multiple-task gait analysis approach: kinematic, kinetic and EMG reference data for healthy young and adult subjects," Gait Posture, vol. 33, no. 1, pp. 6–13, 2011.

[16] M. Wand and T. Schultz, "Pattern learning with deep neural networks in EMG-based speech recognition," in Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE, 2014, pp. 4200–4203.

[17] D. Verstraete, A. Ferrada, E. L. Droguett, V. Meruane, and M. Modarres, "Deep learning enabled fault diagnosis using time-frequency image analysis of rolling element bearings," Shock Vib., vol. 2017, 2017.

[18] L. Yuan and J. Cao, "Patients' EEG Data Analysis via Spectrogram Image with a Convolution Neural Network," in International Conference on Intelligent Decision Technologies, 2017, pp. 13–21.

[19] Z. Xiong, M. K. Stiles, and J. Zhao, "Robust ECG Signal Classification for Detection of Atrial Fibrillation Using a Novel Neural Network," Computing, vol. 44, p. 1, 2017.

[20] A. Vilamala, K. H. Madsen, and L. K. Hansen, "Deep convolutional neural networks for interpretable analysis of EEG sleep stage scoring," ArXiv Prepr. ArXiv171000633, 2017.

[21] O. Tsinalis, P. M. Matthews, Y. Guo, and S. Zafeiriou, "Automatic sleep stage scoring with single-channel EEG using convolutional neural networks," ArXiv Prepr. ArXiv161001683, 2016.

[22] G. Vrbancic and V. Podgorelec, "Automatic Classification of Motor Impairment Neural Disorders from EEG Signals Using Deep Convolutional Neural Networks," Elektron. Ir Elektrotechnika, vol. 24, no. 4, pp. 3–7, 2018.

[23] P. Xia, J. Hu, and Y. Peng, "EMG-Based Estimation of Limb Movement Using Deep Learning With Recurrent Convolutional Neural Networks," Artif. Organs, vol. 42, no. 5, pp. E67–E77, 2018.

[24] "UCI Machine Learning Repository: EMG Physical Action Data Set Data Set." [Online]. Available: https://archive.ics.uci.edu/ml/datasets/EMG+Physical+Action+Data+Set. [Accessed: 02-Dec-2018].

[25] M. P. G. Bhosale and S. T. Patil, "Classification of EEG Signals Using Wavelet Transform and Hybrid Classifier For Parkinson's Disease Detection," Int. J. Eng., vol. 2, no. 1, 2013.

[26] Y. Dong et al., "Evaluations of deep convolutional neural networks for automatic identification of malaria infected cells," in Biomedical & Health Informatics (BHI), 2017 IEEE EMBS International Conference on, 2017, pp. 101–104.

[27] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in Advances in neural information processing systems, 2012, pp. 1097–1105.

[28] Z. Hussain, F. Gimenez, D. Yi, and D. Rubin, "Differential Data Augmentation Techniques for Medical Imaging Classification Tasks," in AMIA Annual Symposium Proceedings, 2017, vol. 2017, p. 979.

[29] M. Gao et al., "Holistic classification of CT attenuation patterns for interstitial lung diseases via deep convolutional neural networks," Comput. Methods Biomech. Biomed. Eng. Imaging Vis., vol. 6, no. 1, pp. 1–6, 2018.

[30] Z. Zhou, J. Y. Shin, L. Zhang, S. R. Gurudu, M. B. Gotway, and J. Liang, "Fine-Tuning Convolutional Neural Networks

for Biomedical Image Analysis: Actively and Incrementally.," in CVPR, 2017, pp. 4761–4772.
[31] M. A. Tanner and W. H. Wong, "The calculation of posterior distributions by data augmentation," J. Am. Stat. Assoc., vol. 82, no. 398, pp. 528–540, 1987.
[32] L. Perez and J. Wang, "The effectiveness of data augmentation in image classification using deep learning," ArXiv Prepr. ArXiv171204621, 2017.
[33] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning (Book in preparation) -. MIT press, 2016.
[34] "Wearable Sensors for Human Activity Monitoring: A Review - Google Scholar." [Online]. Available: https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=Wearable+Sensors+for+Human+Activity+Monitoring%3A+A+Review&btnG=. [Accessed: 28-Nov-2018].

Haya Alaskar is an assistant professor in the College of Computer Science and Engineering at Prince Sattam Bin Abdelaziz University, Saudi Arabia. Haya gained her PhD in Computer Science from Liverpool John Moores University in 2014. She also gained MSc in Applied Artificial Intelligence from University of Exeter in2009. Haya's Research interest include computer applications in Medical and financial application, data science.