Real-Time EMG Signal Classification via Recurrent Neural Networks

Reza Bagherian Azhiri  
Predictive Analytics and Technologies Lab, ME Dept.  
The University of Texas at Dallas  
Richardson, TX, USA  
reza.azhiri@utdallas.edu

Mohammad Esmaeili  
Department of Electrical and Computer Engineering  
The University of Texas at Dallas  
Richardson, TX, USA  
esmaeili@utdallas.edu

Mehrdad Nourani  
Predictive Analytics and Technologies Lab, ECE Dept.  
The University of Texas at Dallas  
Richardson, TX, USA  
nourani@utdallas.edu

Abstract—Real-time classification of Electromyography signals is the most challenging part of controlling a prosthetic hand. Achieving a high classification accuracy of EMG signals in a short delay time is still challenging. Recurrent neural networks (RNNs) are artificial neural network architectures that are appropriate for sequential data such as EMG. In this paper, after extracting features from a hybrid time-frequency domain (discrete Wavelet transform), we utilize a set of recurrent neural network-based architectures to increase the classification accuracy and reduce the prediction delay time. The performances of these architectures are compared and in general outperform other state-of-the-art methods by achieving 96% classification accuracy in 600 msec.

Index Terms—Electromyography, Recurrent Neural Network, Real-Time classification, Deep Learning, Wavelet Transform.

I. INTRODUCTION

Electromyography (EMG)-based pattern recognition of finger movements has been widely accepted by researchers as a promising method for controlling of prosthetic hands. The EMG signal is the result of action potential of muscle tissues after contraction. This transmitted electrical signal is captured by skin-mounted sensors placed on the target muscle. Due to monotonic relation between finger movement and that of its associated EMG signal, the pattern recognition of hand gesture could be detected. This pattern recognition is examined in either offline or online mode. In offline mode, the major goal is to achieve a higher accuracy [1]. The extracted features from the EMG signal and utilized classifier both impact the accuracy. Generally speaking, three major feature sets are utilized for EMG signals classification: in time domain (TD), frequency domain (FD) and time-frequency domain. Various optimization methods such as Particle Swarm Optimization (PSO) and Genetic algorithm (GA) [2], [3] can be employed to select a set of features with significant importance. Also, a plethora number of classifiers have been reported by researchers including decision trees [4], random forest [5], K nearest neighbor (KNN) [6], [7], naïve Bayes classifier [8], multilayer perceptron (MLP) [9], gradient boosting (GB) [10], support vector machine (SVM) [11], [12] and extreme value machine (EVM) [13].

Esa et al. [14] extracted Hudgins features, root mean square (RMS) and combination of all these features and then employed SVM as a classifier to get the accuracy of 86.67% on index finger and 96.67% on thumb finger. Reference [11] performed spectral analysis on EMG signals to extract reflection coefficients as the features and then implemented SVM as a classifier to get 89% accuracy. Azhiri et al. [15] has used the same features but implemented EVM and increased the accuracy to 91%. Authors in [7] extracted fractional fast Fourier transform (FrFT) as their features, and then KNN was applied as a classifier to achieve the accuracy of 98.12%. Bhattachargee et al. [10] extracted RMS, standard deviation, variance and FFT as their features and classified the results by GB classifier with 98.50% accuracy.

While the majority of studies have concentrated on the offline classification of EMG signals, the online performance of system has practical role on the real-time control of prosthetic hands. For such applications, not only the accuracy of a classifier but also the delay time to get that accuracy is important. Also, the accuracy of offline mode generally is more than online mode. Additionally, the accuracy in online systems depends on the delay time in the postprocessing. The less delay time is considered, the less accuracy is achievable.

Khushaba et al. [16] employed Hudgins features, and autoregressive (AR) for feature extraction and applied library SVM (LIBSVM) for classification. They got the average accuracy of 90% in 800 msec. In another online classification approach [17], wavelet packet transform as a generalized format of wavelet transform has been implemented on the EMG signals for feature extraction. Then by employing principal component analysis (PCA) and self-organizing feature map (SOFM), the number of features has been reduced. Finally, a multilayer perceptron (MLP) was used as a classifier. Jaramillo et al. [9] applied filtering and rectifying on original signals, extracted features in time, frequency and time-frequency
domains, and utilized various parametric and nonparametric classifiers.

Recently, multifarious deep-learning based classifiers have been utilized in EMG-based hand gesture classification in online mode. Deep learning has the advantage of stability which reduces the errors associated with environmental noises especially in real-time EMG signal classification [18].

In classification with recurrent neural network (RNN), Nasri et al [19] achieved the accuracy of 77.85% on EMG dataset including 6 gestures with RNN structure. In order to enhance the accuracy of classification, Koch et al. [20] introduced a new loss function on the outputs of RNN in which true predictions have more weights. The accuracy for true predictions has improved by 10%. Reference [21] compared different structures including feed-forward neural network (FFNN) as a static model and RNN, long-short term memory (LSTM) and gated recurrent units (GRU) as dynamic models to examine the temporal information of EMG signals. Using gesture detection accuracy as a criteria for the evaluation of proposed model, they concluded that both static and dynamic models have similar accuracies. Also, it seems using other techniques like ensemble learning [?] and other neural network architectures such as graph convolutional neural network (GCNN) [22], to make connection between different sensors, can significantly improve the classification accuracy.

In this paper, we investigate recurrent neural network-based architectures for real-time classification of EMG signals. Discrete wavelet decomposition is employed to extract the features from time-frequency domain to feed into the recurrent neural networks. These architectures achieve a better classification accuracy in using shorter time in comparison with the state-of-the-art methods.

II. PROPOSED ARCHITECTURE

Our proposed architecture, as shown in Fig. 1, includes three main steps: preprocessing, processing, and postprocessing.

A. Preprocessing

EMG signals have a stochastic behavior such that instantaneous processing is unable to generate favorite information for real-time classification of EMG signals. The input data is divided into the batches of consecutive samples called data windowing. Each window has the size of 400 samples with 200 overlapped samples with the previous window. Overlapped windows generally have better classification results than disjointed ones. The wavelet transform features are calculated for each window and used as inputs to the RNN structure. In this paper, we consider a 2-level db1 mother wavelet transform decomposition.

B. Recurrent Neural Network Structures

Recurrent neural networks (RNNs) are part of artificial neural networks that are appropriate to exhibit temporal dynamic behavior. RNNs are able to use their memory to process variable length sequences of inputs. This feature makes RNNs a valuable competitor against NNs and CNNs. RNNs have applications such as speech recognition and handwriting recognition, in dealing with sequence data.

Now we introduce two basic units (BUs) that are employed in our architectures. Fig. 3 illustrates two basic units BU1 and BU2.

- **BU1**: consider \( W_i, b_i \) and \( w_a \) as the parameters of the basic unit 1. This basic unit takes vectors \( x \) and \( a_p \) as its inputs and returns two outputs including \( \hat{y} = A_3 \) and
Fig. 4. The recurrent neural network architecture used in this paper.

Fig. 5. The bidirectional recurrent neural network (BRNN) architecture used in this paper.

Fig. 6. Fifteen finger movements classes for individual and combined fingers [23] Each finger movement has related to a class as: (1. Thumb, 2. Index, 3. Middle, 4. Ring, 5. Little, 6. Thumb-Index, 7. Thumb-Middle, 8. Thumb-Ring, 9. Thumb-Little, 10. Index-Middle (I-M), 11. Middle-Ring (M-R), 12. Ring-Little (RL), 13. Index-Middle-Ring (I-M-R), 14. Middle-Ring-Little (M-RL), and 15. close hands (HC).)

\[ a_n = Z_3 \] where

\[ A_0 = x \]
\[ Z_{i+1} = W_i A_i + b_i \quad \text{for} \quad i = 0, 1, 2 \]
\[ A_{i+1} = \tanh(Z_{i+1} + w_a p) \quad \text{for} \quad i = 0 \]
\[ A_{i+1} = \tanh(Z_{i+1}) \quad \text{for} \quad i = 1 \]
\[ A_{i+1} = \text{Softmax}(Z_{i+1}) \quad \text{for} \quad i = 2. \]

and \( a_n = a_p = Z_3 \) where

\[ A_{i+1} = \tanh(Z_{i+1} + w_a a_p + w_{\hat{a}} \hat{a}_p) \quad \text{for} \quad i = 0. \]

We construct an original recurrent neural network (RNN) and a bidirectional recurrent neural network (BRNN) by stacking a set of the basic units BU1 and BU2, respectively. The number of basic units that are used at each architecture is a hyperparameter which is tuned empirically. In this paper, five basic units are employed for each architecture. Also, we consider two different input approaches for each architecture as follows:

- BU2: consider \( W_i, b_i, w_a \) and \( w_{\hat{a}} \) as the parameters of the basic unit 1. This basic unit takes vectors \( x, a_p \) and \( \hat{a}_p \) as its inputs and returns three outputs including \( \hat{y} = A_3 \)

- Same inputs: in this approach, all the basic unit inputs
are the same, i.e.,
\[ x_1 = x_2 = x_3 = x_4 = x_5 = f(w_i) \]
where \( f(w_i) \) refers to the extracted features of window \( w_i \) from the original signal. In this method, since all the basic unit inputs are the same, the RNN structure tries to improve the uncertainty at each basic unit during the training phase.

- Sequential inputs: in this approach, the basic unit inputs are different. Indeed, the input of each basic unit is the extracted features from a shifted window (50 msec) of the original signal, i.e.,
\[ x_j = f(w_{i+j-1}) \text{ for } j = 1, \cdots, 5. \]
In this method, since the basic unit inputs are different, the RNN structure tries to learn the relationship between the inputs to reduce the uncertainty during the training phase [24].

C. Postprocessing

To refine the results of RNN structure, a postprocessing technique after classification of RNN structure is necessary. Each window performs a class decision, and postprocessing step minimizes ambiguities and false misclassifications [16]. In fact, postprocessing step helps us to combine the results gained by each window of the EMG signal. In this paper, the majority voting approach is selected in which the elements of each class is counted and a class as identified by majority is chosen.

III. FEATURE EXTRACTION

In the time domain, feature extraction functions are directly applied on a window of raw signals. Feature extraction functions are able to extract statistical features from the raw signal. There are several features that are commonly used in the literature. In this paper, we will use 19 of these features and their corresponding functions. Table 1 summarizes the mathematical expressions of these functions.

Discrete wavelet decomposition (DWD) provides sufficient information both for analysis and synthesis of the original signal, with a significant reduction in the computation time. Discrete wavelet transform can be implemented with a single level or multiple levels. A two-level discrete wavelet decomposition is shown in Fig. 2.

The discrete wavelet transform of a signal is calculated by passing it through a series of filters. The samples are passed through a low pass filter. The signal is also decomposed simultaneously using a high-pass filter. The outputs give the detail coefficients (from the high-pass filter) and approximation coefficients (from the low-pass filter). It is important to note that these two filters are related to each other and they are known as a quadrature mirror filter. The decomposition is repeated to further increase the frequency resolution and the approximation coefficients decomposed with the high-pass and the low-pass filters and then down-sampled. After passing a window of raw signal through the high-pass and low-pass filters, the feature extraction functions in Table II are applied on the detail coefficients and approximation coefficients. In this paper, we consider a 2-level db1 mother wavelet transform decomposition. Therefore, for a two-level wavelet decomposition, we will extract 57 features (3 layers of 19 wavelet features) for each window of the raw signal.

IV. EXPERIMENTAL RESULTS

A. Dataset

The datasets which are used in this paper are provided by Center of Intelligent Mechatronic Systems at the University of Technology at Sydney [16]. The datasets include EMG data of eight participants (six men and two women in range of 20 to 30 years old). EMG signals are sensitive to the condition of experiment and effects of neighbor limbs. For uniformity, participants sat on the armchair such that their arms were fixed and stable. To firmly stick the sensors to the skin above the targeted muscle, an adhesive skin interface was used. Resulting EMG signals from the electrodes were amplified to get the total gain of 1000. The data was recorded at 4,000 Hz applying an analog-to-digital converter, a bandpass filter between 20 and 40 Hz and a notch filter to remove the 50 Hz line interference. This filtering is necessary to remove noises resulting by motion of artifacts and high-frequency random noise. Participants performed each movement of fingers six times while the duration of each movement from rest pose to contraction pose is five seconds.

In the first dataset, participants performed ten different finger movements consisting five individuals shown as thumb (T), index (I), middle (M), ring (R), little (L) and five combined movements named as thumb-index (T-I), thumb-middle (T-M), thumb-ring (T-R), thumb-little (T-L) and closed hands using two-channel (2C) sensors. We select the first four trials for the training of the classification methods and the last two for the test of the methods in order to check the accuracy of the classifiers. Each gesture took five seconds including rest and holding of each finger posture.

Second dataset includes total of 15 classes: five individuals finger movement thumb (T), index (I), middle (M), ring (R), little (L) and ten combined finger movements including thumb-index (T-I), thumb-middle (T-M), thumb-ring (T-R), thumb-little (TL), index-middle (I-M), middle-ring (M-R), ring-little (RL), index-middle-ring (I-M-R), middle-ring-little (M-RL), and close hands (HC) using eight-channel (8C) sensors. Each gesture took 20 seconds including rest and holding of each finger gesture. The first two trials were selected for the training of the model and the last one was used for the test to check the accuracy of proposed classifiers. Fig. 6 depicts these datasets that first ten finger movements are considered in the first dataset and entire fifteen movements are considered in the second dataset. The details of dataset, number of subjects and condition of experiments are reported in Table II.

B. Results and Discussion

In this section, the classification results of RNN and bidirectional RNN (BRNN) are presented and compared. Table III
one-sided RNN architecture. Indeed, sharing the features that is able to capture the relationship of inputs better than a two-sided connectivity of basic units, the BRNN architecture among all other proposed architectures of RNN. Due to the higher accuracy in results show that BRNN with the same inputs could achieve which is used for the postprocessing is majority voting. The compares the test accuracy of proposed RNN architectures at different signal lengths for the first dataset. The method which is used for the postprocessing is majority voting. The results show that BRNN with the same inputs could achieve the higher accuracy in 600 msec which is the best performance among all other proposed architectures of RNN. Due to the two-sided connectivity of basic units, the BRNN architecture is able to capture the relationship of inputs better than a one-sided RNN architecture. Indeed, sharing the features that are learned by each basic unit reduces the misclassification and improves overall accuracy. Also, considering the same inputs for all basic units enhances the classifier performance by providing a prior information about the inputs. In other words, the uncertainty that exist between the basic units is reduced in BRNN compared to RNN.

Again, various architectures of RNN at different signal lengths are compared for eight-channel (8C) dataset. BRNN with the same inputs and RNN with the sequential inputs show better performance than other architectures. Both architectures with the same inputs and RNN with the sequential inputs show better performance than other architectures. Both architectures achieve the accuracy of 93.3% in 500 msec. These results are summarized in Table IV

In Fig. 7 and Fig. 8, the classification accuracy of 2C and 8C datasets for BRNN with the same inputs are shown. For the 2C dataset at the 600 msec signal length, the accuracy of ring, thumb-index, thumb-middle and thumb-little is 100%. For 8C dataset at the 500 msec signal length, the accuracy for thumb, index, middle, little, thumb-middle, thumb-ring, middle-ring and index-middle-ring is 100%. Comparing the accuracies of classes, a uniform accuracy for each finger movement in both datasets is observed. Such uniform accuracy for different classes is important factor that proves the proposed BRNN architecture has stable performance on various finger
TABLE III
THE ACCURACY (IN %) OF DIFFERENT RNN STRUCTURES FOR VARIOUS SIGNAL LENGTH FOR 2C DATASET.

| RNN Structures                  | Signal length (msec) |
|---------------------------------|----------------------|
|                                 | 100  | 150  | 200  | 250  | 300  | 350  | 400  | 450  | 500  | 550  | 600  |
| RNN with the same inputs        | 62.0 | 62.0 | 72.5 | 74.0 | 78.0 | 79.5 | 84.5 | 86.0 | 85.5 | 89.0 | 90.0 |
| BRNN with the same inputs       | 70.5 | 70.5 | 74.0 | 81.5 | 87.0 | 90.0 | 91.0 | 92.5 | 92.0 | 91.5 | 96.0 |
| RNN with sequential inputs     | -    | -    | -    | -    | 85.5 | 85.5 | 88.0 | 89.0 | 91.0 | 93.0 | 93.0 |
| BRNN with sequential inputs    | -    | -    | -    | -    | 85.0 | 85.0 | 91.0 | 91.5 | 93.5 | 93.5 | 93.5 |

TABLE IV
THE ACCURACY (IN %) OF DIFFERENT RNN STRUCTURES FOR VARIOUS SIGNAL LENGTH FOR 8C DATASET.

| RNN Structures                  | Signal length (msec) |
|---------------------------------|----------------------|
|                                 | 100  | 150  | 200  | 250  | 300  | 350  | 400  | 450  | 500  | 550  | 600  |
| RNN with the same inputs        | 82.5 | 82.5 | 81.6 | 84.1 | 87.5 | 84.1 | 86.6 | 87.5 | 90.8 | 91.6 | 91.6 |
| BRNN with the same inputs       | 85.0 | 85.0 | 90.0 | 90.8 | 92.5 | 92.5 | 92.5 | 92.5 | 93.3 | 93.3 | 93.3 |
| RNN with sequential inputs     | -    | -    | -    | -    | 89.1 | 89.1 | 90.8 | 90.8 | 90.8 | 91.6 | 92.5 |
| BRNN with sequential inputs    | -    | -    | -    | -    | 89.2 | 89.2 | 90.8 | 90.8 | 93.3 | 93.3 | 93.3 |

TABLE V
COMPARING THE RESULTS OF DIFFERENT APPROACHES FOR THE SAME (2C) EMG DATASET [16].

| Methods                          | Features                     | Classifier | Classes | Signal length (msec) | Avg. Acc (in %) |
|----------------------------------|------------------------------|------------|---------|----------------------|-----------------|
| [30]                             | TD+Hjorth+RMS                | ANN        | 5       | 5000                 | 96.7            |
| [31]                             | AR + RMS                     | ANN+NMF    | 5       | 5000                 | 92              |
| [11]                             | Reflection Coefficients      | SVM        | 10      | 5000                 | 89              |
| [32]                             | FNPA                         | ELM+libSVM+RegTree | 10      | 5000                 | 91              |
| [15]                             | Reflection Coefficients      | EVM        | 10      | 5000                 | 91              |
| [33]                             | Mixture of Features          | Random Forest | 10      | 5000                 | 93.75           |
| [14]                             | Hudgind+RMS                  | SVM        | 10      | 5000                 | 96.67 (just thumb finger) |
| [34]                             | FrFT                         | KNN        | 10      | 5000                 | 98.12           |
| [10]                             | statistics features+FFT      | GB         | 10      | 5000                 | 98.5            |
| [35]                             | Mixture of Features          | SVM        | 10      | 5000                 | 96.5            |
| [16]                             | TD+AR+Hjorth                 | KNN+SVM+Fusion | 10      | 800                  | 90              |
| [15]                             | Wavelet                      | Deep Neural Network | 10      | 800                  | 95.5            |

Ours Wavelet BRNN 10 600 96

movements.

The accuracy of classification depends on several factors including the number of classes, number of sensors, extracted features and utilized classifier. Increasing the number of classes will decrease the accuracy of classifications. Also, the number of sensors used for the experiments can significantly affect the final result. Therefore, we compare our results with other researchers with the same situation on the same dataset. In Table V, the results of 2C EMG dataset are compared, while in Table VI the 8C EMG dataset is considered. The offline results have generally better accuracies but as it is discussed, for real world applications and motion control of prosthetics, online classification have practical role. Rare researches have been done on the online classification of EMG signals. Our result is very competitive such that the accuracy and signal length are simultaneously regarded and better accuracy in less signal length is gained for both datasets.

We have trained this model in python and ran it in a windows-based PC with 2.60 GHz CPU and 16 GB memory. The delay time for the system is the result of feature extraction and process time required by the classifier to make the decision. For 2C dataset using BRNN at 600 msec signal length at postprocessing, delay time is about 4.72 msec for the feature extraction and 3.12 msec for the classification.
TABLE VI
COMPARING THE RESULTS OF DIFFERENT APPROACHES FOR THE SAME (8C) EMG DATASET [23].

| Methods        | Features                     | Classifier                  | Classes | Signal length (msec) | Avg. Acc (in %) |
|----------------|------------------------------|-----------------------------|---------|----------------------|-----------------|
| [23]           | MCA                          | SVM+KNN+ELM                 | 15      | 20000                | 95.0            |
| [36]           | Mixture of Features          | NB+KNN+MLP+QDA+SVM+ELM      | 15      | 20000                | 90              |
| [37]           | WPT-4+TD+AR-4+RMS            | QDA+KNN+SVM                 | 15      | 20000                | 98.5            |
| [38]           | Raw Data                     | CNN                         | 15      | 100                  | 91.26           |
| **Ours**       | Wavelet                      | BRNN                        | **15**  | **500**              | **93.33**       |

Fig. 7. The classification accuracy for each of 10 classes of BRNN for 2C dataset at signal length of 600 msec.

Fig. 8. The classification accuracy for each of 15 classes of BRNN for 8C dataset at signal length of 500 msec.

V. CONCLUSION

To overcome the challenges of real-time classification of EMG signals, we introduced two new recurrent neural network-based architectures. Discrete wavelet transformation was utilized to extract the features from time-frequency domain to feed into the recurrent neural networks. Each architecture was investigated for two different input forms. The performance of each architecture was evaluated and it was compared with the state-of-the-art approaches. It was shown that at least one of these architectures increases the classification accuracy, reduces the delay time, and outperforms other architectures.

REFERENCES

[1] S. Rouhani, T. Rahman, and V. Gogate, “Algorithms for the nearest assignment problem,” in *IJCAI*, pp. 5096–5102, 2018.
[2] M. Esmaeili, N. Granpayeh, and M. Bozorgi, “A novel reliable optimization method for output beam forming of photonic crystal waveguide terminated with surface crow,” *Optik*, vol. 126, no. 4, pp. 421–425, 2015.
[3] A. A. M. Lima, R. M. Araújo, F. A. G. d. Santos, V. H. Yoshizumi, F. K. H. de Barros, D. H. Spatti, L. H. B. Liboni, and M. E. Dajer, “Classification of hand movements from emg signals using optimized mlp,” in *2018 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–7, 2018.
