EMG based Hand Gesture Classification using Empirical Mode Decomposition Time-Series and Deep Learning

Deniz Hande KISA, Mehmet Akif OZDEMIR, and Onan GUREN
Department of Biomedical Engineering
Izmir Katip Celebi University
Izmir, TURKEY
denizhandekisa@hotmail.com, makif.ozdemir@ikcu.edu.tr, onan.guren@ikcu.edu.tr

Aydin AKAN
Dept. of Electrical and Electronics Eng.
Izmir University of Economics
Izmir, TURKEY
akan.aydin@ieu.edu.tr

***Abstract*—Computer systems working with artificial intelligence can recognize movements and gestures to be used for many purposes. In order to perform recognition, the electrical activity of the muscles can be utilized which is represented by electromyography (EMG) and EMG is not a stationary biological signal. EMG based movement recognition systems have an important place in distinct areas like in human-computer interactions, virtual reality, prosthesis, and hand exoskeletons. In this study, a new approach based on deep learning (DL) and Empirical Mode Decomposition (EMD) is proposed to improve the accuracy rate for recognition of hand movements in its application areas. Firstly, 4-channel surface EMG (sEMG) signals were measured while simulating 7 different hand gestures, which are extension, flexion, ulnar deviation, radial deviation, punch, open hand, and rest, from 30 subjects. After that, noiseless signals were procured utilizing filters as a result of preprocessing. Then, pre-processed signals were subjected to segmentation. Thereafter, the EMD process was applied to each segmented signal and Intrinsic Mode Functions (IMFs) were obtained. The IMFs time-series which are some kind of screen images of the first 3 IMFs have been recorded. For classification, IMFs images have given as inputs and have trained to the 101-layer Convolution Neural Network (CNN) based on Residual Networks (ResNet) architecture, which is a DL model.**

Keywords—Convolutional Neural Network (CNN), Deep Learning, Electromyography (EMG), Empirical Mode Decomposition (EMD), Hand Gesture, Intrinsic Mode Function (IMF), ResNet.

I. INTRODUCTION

In the movement system, the human body can move with the harmonic acting mechanisms of the bones, joints, and muscles. The ability to move is an imperative factor in human daily lives from communication to actions. The most active component of that system is the muscle and it gives information about movement [1]. For this purpose, EMG is utilized to investigate the working mechanism of muscles and their effects on the movement. EMG does not have a symmetrical structure and it is nonstationary bio-signal as a sum of biopotentials during contraction. It is assumed that in EMG, there is knowledge about which movement takes place. These signals can be taken with a sEMG electrode that is a major and noninvasive technique. For control strategy of rehabilitation and myoelectric based devices, EMG signals can be utilized, and also, they need to be classified [1].

The capability to recognize the hand movements of a computer system or device may be utilized in many potential applications which are sign language recognition, robotics, virtual reality, and human-computer interaction (HCI). Especially, HCI is significant in military and medical areas that include studies of hand gestures such as real-time controlling with sEMG for the prosthesis of individuals, the haptic systems, and the exoskeletons [2].

In myoelectric controlled based systems, to ensure movement prediction and recognition, EMG signals need to be processed and classified. It has happened essentially to make the classification of EMG signal utilized for control especially in the multi-grip or upper limb prosthesis. EMG gives information about movement type and where it takes place systems [3]. Hand gesture recognition can be utilized to predict the movement in exoskeleton systems to provide more advanced movement prediction to obtain better synchronization [4]. sEMG based pattern recognition systems have shown good potency for controlling the described systems [3].

In order to procure better HCI synchronization, there need new techniques to improve the accuracy percentage and the synchronization capability in myoelectrical control systems in movement recognition, particularly in the recognition of hand gestures. Many engineering and medical practices for EMG-based hand gesture recognition become available [4]. Before classification and recognition, to analyze the signals, useful methods are needed. For this aim, EMD can be utilized. EMD is an analyzing method for non-stationary and nonlinear time-series and decomposes the signal into a sequence of swings named IMFs. By utilizing distinct time and frequency domain technics, the multichannel biological signal's IMFs extracted by EMD. It increases the accuracy rate and demonstrates good

This work was supported by the Scientific and Technical Research Council of Turkey (TUBITAK) under Grant No. 1919B011903429.

978-1-7281-8073-1/20/$31.00 ©2020 IEEE

performance than complex and traditional preprocessing methods [5]. The EMD method is the process of shredding a signal without leaving the time domain. EMD is useful for any application that requires filtering EMG signals during the pre-processing stage. Previous studies have shown that it can be applied successfully to decrease EMG noise. The main distinction of the method is that it performs signal decomposition adaptively and only based on available data and uses a prefixed filter set [6]. Additionally, numerous DL-based hand gesture recognition works have been published. Recent DL architectures propound high accuracies (>95%) [7]. So, to obtain high accuracy, CNN architecture as the DL method was utilized in this study.

Sapsanis *et al.* presented a pattern recognition method to recognize hand movements utilizing sEMG data. They decomposed EMG signals utilizing EMD into IMFs and then features were extracted. They tested their results and the results committed that EMD could enhance the distinctive capability of the traditional feature series obtained from raw EMG. For instance, the average accuracy increased from the raw EMG extracted features of 86.92% to all extracted features 90.42% for a subject [8].

Yan *et al.* presented an effective and efficient combination of feature extraction and multiclass classifier for motion classification by analyzing sEMG signals. They introduced EMD to decompose the EMG signals into a few IMFs and after that calculated the coefficients of each IMF to effectuate feature set. They designed the multi-class classifier based on least squares support vector machines (LS-SVMs) for the classification of varied movements. Their results demonstrated that the accuracy of movement recognition was developed [9].

Xiaojing *et al.* investigated feature extraction and classification of sEMG signals. Firstly, they used an independent component analysis technique to remove the power frequency interference, and then the processing of low noise signal was performed by EMD. After that, the decomposed signal was utilized to establish the autoregressive model. They utilized coefficients of the model qua features of signal and optimized probabilistic neural network for classification of 6 forearm movements. Their results showed that the proposed method was effective for extraction and classification [10].

In this study, collected EMG signals are segmented and IMFs are created via the EMD process and recorded the time-series IMFs, which are some kind of screen images of the first 3 IMFs. These images of IMFs are given as input data in the CNN structure which is ResNet-101.

II. Materials and Method

A. EMG Dataset

In this study, sEMG signals utilized for hand gesture recognition were recorded by the MP36 model BIOPAC device with 4-channels. 30 healthy volunteers (15 females and 15 males) take place in this study. Their ages were between 18 to 25. The sEMG signals were recorded at 2 kHz sampling frequency for 7 different hand gestures utilizing Ag/AgCl surface bipolar electrodes. The seven hand gestures (in Fig. 2) are extension, flexion, ulnar deviation, and radial deviation of the wrist, punch, open hand, and rest position. The recorded EMG signals' amplitude is between 0-10 mV or 0-1.5 mV.

To collect data without noise, it has purposed to take data from surface muscles that were closer to skin. The EMG signals have usually minor amplitudes. When measured muscle's number improves, the biological signal's number that can be measured and monitored enhances. According to this, 4 distinct surface muscles have been selected that the beneficial and non-noisy signals can be received pending the hand gestures.

The utilized 4 muscles are flexor carpi radialis, flexor carpi ulnaris, extensor carpi radialis, and extensor carpi ulnaris. After cleaning of the skin by alcohol to remove dead cells and oils, sEMG electrodes are located on their approximate location shown in Fig. 1. The sEMG recording time is taken 515 seconds which included 5 cycles.

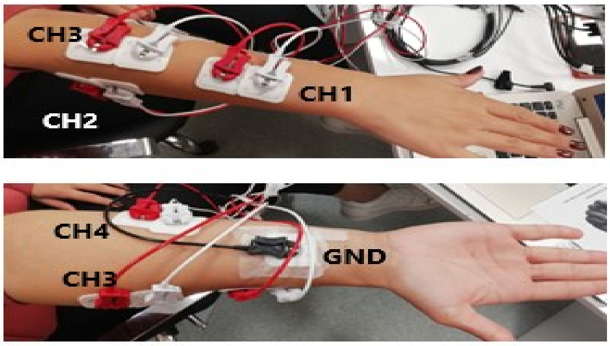

Fig. 1. Four channels of sEMG electrodes placement.

B. Pre-Signal Processing and Segmentation

The filtering process was performed and the raw sEMG signals from volunteers were cleaned from noises. Then noises, which were caused by external sources or body, during muscle contractions, were filtered with a digital low pass filter between 5-500 Hz and a 50 Hz Notch filter. After filtering, segmentation was performed and 4-channel sEMG signals were divided into 4-seconds time windows correspondent the instants when the gesture acting. This process enhanced the number of data needed for DL and additionally, it provided not to miss valued knowledge at the duration of gesture in the signal. An example of segmented EMG signals pertains to hand gestures can be seen in Fig. 2. 5 cycles were repeated and performed by each participant. For these reasons, total 4-channel sEMG segments appertain to each hand gesture can be calculated as; the *number of time-series images = 30 participants x 5 reps x 7 hand gestures x 4 channels x 4 IMFs (one of the original signals' time-series) = 16800.*

C. Empirical Mode Decomposition of sEMG Segments

After segmentation, the EMD method was utilized to each segmented signal then IMFs were obtained and recorded the time-series IMFs, which are the time-series snap screens for the first 3 IMFs.

sEMG signal is a nonstationary and nonlinear like other signals in nature. So, utilizing algorithms that consider linearity and stationarity features may be improper. At this point, EMD ensures a new adaptable method nonstationary signal for their analysis [8].

EMD behaves like a non-linear and adaptive filter, it decomposes the signal into several IMFs. An IMF symbolizes a basic oscillating function fulfilling two circumstances [8]:

i. The zero passing's number and the local extreme's number are the same or different by one.
ii. The local mean is equal to zero, and it is described with the mean of local maximum and local minimum.

The two circumstances engage that all minima of an IMF are negative and all its maxima are positive [8]. The signal **x(t)** and the EMD algorithm can be abstracted as followings:

i. Adjusting the whole local maximum and a local minimum of **x(t)** signal.
ii. Interpolating between consecutive the local maximum and minimum via cubic function and forming an upper envelope ($\mathbf{e_{max}(t)}$) and a lower ($\mathbf{e_{min}(t)}$) envelope [11].
iii. Calculating the average of upper and lower envelopes.

$$m(t) = \frac{e_{min}(t) + e_{max}(t)}{2} \tag{1}$$

iv. Subtracting the average from **x(t)** signal to take the detail.

$$d(t) = x(t) - m(t) \tag{2}$$

v. If the number of local extremes of **d(t)**, is equal to or different from the number of zero crossings by one, and the mean of **d(t)** is close to zero, then $\mathbf{IMF_1} = \mathbf{d(t)}$**.** Else, repeating steps **i** to **iv** on **d(t)** in place of x(t), till recent **d(t)** supplies the circumstances of an IMF in step **v**.
vi. Calculating residue $r(t) = x(t) - d(t)$.
vii. If **r(t)** is over the threshold, then repeating steps **i** to vi on **r(t)** is performed to procure following IMF and novel residue.

Finally, **n** vertical IMFs are procured off which the initial **x(t)** signal can be defined as following [12]. Whence, finally decomposition of initial **x(t)** signal is performed and decomposed into an aggregate of IMFs plus a residual premises [8]:

$$x(t) = \sum_i IMF_i(t) + r(t) \tag{3}$$

Hereafter, IMFs are named as first-order IMFs that derived utilizing the ordinary EMD process. The highest frequency oscillations in the initial **x(t)** signal are represented by the first IMF. The following IMFs cover inferior frequency oscillations of **x(t)**. The last residue demonstrates just overall propensities of the signal [12].

D. *Deep Learning Architecture*

The first 3 IMFs, which are time-series snap screens from EMD, were given as inputs and have trained to the 101-layer CNN for classification. The working principle of CNN is based on the visual cortex and it is designed to imitate the linkage model of neurons in the brain. CNN contains three types of layers, which are convolutional, pooling, and fully connected. In CNN, as the layer's number scale up, the network happens difficult for training, and the accuracy arrives satiation and then goes to reduce. Residual learning lends assistance to resolve that decreasing accuracy issue. Residual learning uses quick path linkages qua a training technic to directly assign the input not simply to the next adjacent one but withal to another subsequent layer, for the network's training. It can be readily explicated as the extraction of input characteristics learned off that layer and it is performed by ResNet utilizing bypass linkages to every couple of 33 filters. In this way, the issue of vanishing gradients is kept away by reusing activation from the previous layer.

In this study, 101-layers ResNet-101 was utilized to train by the time-series images of IMFs obtained from sEMG signals recorded pending 7 distinct hand gestures. Its network shares 4 types of residual learning blocks. ResNet-101 is obtained by modifying ResNet-50. In this study, residual network-based CNN architecture is used to prevent partially similar IMF time-series snap screens from going to overfitting.

III. Results

In this study, a total of 4200 time-series images for every group, which obtained from the original signals, IMF_1, IMF_2, and IMF_3, were created to train the network and given to feed the ResNet-101 network. For every training process total of 4200 images were used to train and this was repeated 4 times. These images were reserved by the validation split method, 80% for the training of the network and the remaining 20% for testing the trained model. The training results validation and test accuracy graphs and the confusion matrixes for all test groups are shown in Fig 3. When the results obtained from the original signal were examined it was seen that loss value was calculated as 0.0862, training accuracy as 95.92%, and validation loss as 0.3019. The validation accuracy of the original signal was 94.22%. F1 score was calculated as 79.67% and mean squared error was found as 0.322619.

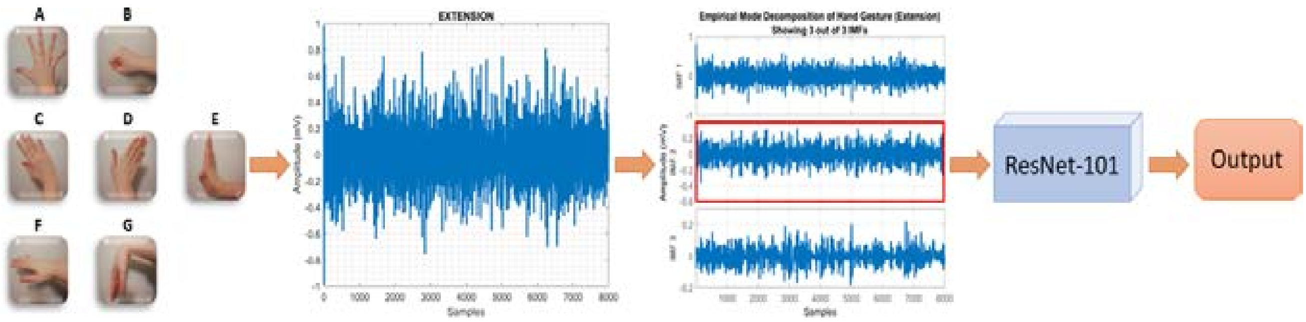

Fig. 2. Block representation of this study; seven different hand gestures, a) Open Hand, b) Punch, c) Radial Deviation, d) Ulnar Deviation, e) Extension, f) Rest, and g) Flexion; an example of segmented sEMG drawings of extension and its EMD time-series output of first three IMFs; ResNet block; and the SoftMax output representation (The sEMG data used in this figure belongs to the 1st sEMG channel of the number #4 participant in the database).

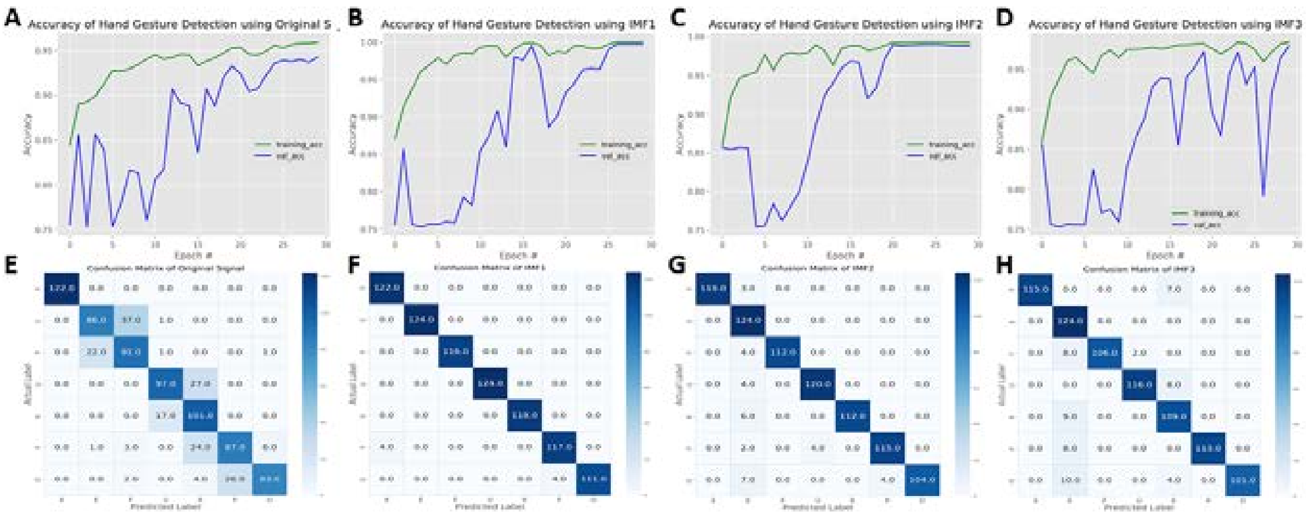

Fig. 3. Network training results, training and validation accuracy for a) using only original signals' time-series snap screens, b) IMF_1 snap screens, c) IMF_2 snap screens, d) IMF_3 snap screens, and confusion matrixes for e) using only original signals, f) IMF_1, g) IMF_2, h) IMF_3, (E: Extension, F: Flexion, O: Open Hand, P: Punch, R: Radial Deviation, X: Rest, and U: Ulnar Deviation).

For the training of IMF_1, it was calculated as 3.3257e-05, training accuracy value as 100%, and validation loss as 0.0088. The validation accuracy of IMF_1 was 99.73%. Its F1 score was calculated as 99.05%, and mean squared error was found as 0.123810. For the training of IMF_2, loss value was calculated as 0.0179, training accuracy value as 99.28%, and validation loss as 0.0312. The validation accuracy of IMF_2 was 98.86%. Its F1 score was calculated as 96.05% and mean squared error was found as 0.361905. For the training of IMF_3, training loss was calculated as 0.0398, accuracy value as 98.42%, and validation loss as 0.0563. The validation accuracy of IMF_3 was 97.94%. Its F1 score was calculated as 93.54% and mean squared error was found as 0.720238.

IV. Conclusion

In this study, IMFs of sEMG signals obtained from 7 different hand gestures were created via EMD. Screenshots of the first 3 IMFs were used to train the ResNet-101 architecture. To compare the success of IMFs, screenshots of original sEMG signals were trained in CNN based architecture.

When the training results are examined, we can say that all training results have achieved very effective accuracy in classifying hand movements. Besides, time-series images obtained from IMF_1 are more successful than other groups. Moreover, the original signal gave the worst results and it further proves the success of IMFs by EMD. Also, the reason why this study provides accurate results as mentioned is that the idea of processing time snaps in the DL model is simple and effective.

As a conclusion, EMD is a useful and feasible process to apply sEMG signals, which shows the nonstationary and nonlinear structure, and this application provides to obtain more information from these signals and obtain better accuracy rate results than raw or original sEMG for recognition of hand movements. In future works, by using larger datasets and distinct hand gestures the proposed method can be developed. In this way, the applicability of the method can be increased.

References

[1] M. Y. Iscan and M. Steyn, *The human skeleton in forensic medicine*. Charles C Thomas Publisher, 2013.

[2] M. J. Cheok, Z. Omar, and M. H. Jaward, "A review of hand gesture and sign language recognition techniques," *International Journal of Machine Learning and Cybernetics,* vol. 10, no. 1, pp. 131-153, 2019.

[3] A. K. Mukhopadhyay and S. Samui, "An experimental study on upper limb position invariant EMG signal classification based on deep neural network," *Biomedical Signal Processing and Control,* vol. 55, p. 101669, 2020.

[4] J. L. Ren, Y. H. Chien, E. Y. Chia, L. C. Fu, and J. S. Lai, "Deep Learning based Motion Prediction for Exoskeleton Robot Control in Upper Limb Rehabilitation," in *2019 International Conference on Robotics and Automation (ICRA)*, 2019, pp. 5076-5082: IEEE.

[5] E. Izci, M. A. Ozdemir, R. Sadighzadeh, and A. Akan, "Arrhythmia detection on ECG signals by using empirical mode decomposition," in *2018 Medical Technologies National Congress (TIPTEKNO)*, 2018, pp. 1-4: IEEE.

[6] A. O. Andrade, S. Nasuto, P. Kyberd, C. M. Sweeney-Reed, and F. Van Kanijn, "EMG signal filtering based on empirical mode decomposition," *Biomedical Signal Processing and Control,* vol. 1, no. 1, pp. 44-55, 2006.

[7] G. Jia, H. K. Lam, J. Liao, and R. Wang, "Classification of Electromyographic Hand Gesture Signals using Machine Learning Techniques," *Neurocomputing,* 2020.

[8] C. Sapsanis, G. Georgoulas, A. Tzes, and D. Lymberopoulos, "Improving EMG based classification of basic hand movements using EMD," in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2013, pp. 5754-5757: IEEE.

[9] Z. G. Yan, Z. Z. Wang, and X. M. Ren, "Joint application of feature extraction based on EMD-AR strategy and multi-class classifier based on LS-SVM in EMG motion classification," *Journal of Zhejiang University-SCIENCE A,* vol. 8, no. 8, pp. 1246-1255, 2007.

[10] S. Xiaojing, T. Yantao, and L. Yang, "Feature extraction and classification of sEMG based on ICA and EMD decomposition of AR model," in *2011 International Conference on Electronics, Communications and Control (ICECC)*, 2011,pp.1464-1467: IEEE.

[11] S. Yol, M. A. Ozdemir, A. Akan, and L. F. Chaparro, "Detection of epileptic seizures by the analysis of EEG signals using empirical mode decomposition," in *2018 Medical Technologies National Congress (TIPTEKNO)*, 2018, pp. 1-4: IEEE.

[12] R. Damaševičius, M. Vasiljevas, I. Martišius, V. Jusas, D. Birvinskas, and M. Wozniak, "BoostEMD: an extension of EMD method and its application for denoising of EMG signals," *Elektronika ir elektrotechnika,* vol. 21, no. 6, pp. 57-61, 2015.