EMG Signal Classification with Effective Features for Diagnosis

Abdul Wadud
Department of Computer Science
South Asian University
New Delhi, India
Email: dewanwadud.ex@gmail.com

Md. Imran Hossain Showrov
Department of Computer Science
South Asian University
New Delhi, India
Email: showrov.cse@gmail.com

***Abstract*—Electromyography (EMG) signals are broadly used in various clinical or biomedical applications, prosthesis or rehabilitation devices, Muscle-Computer Interface (MCI), Evolvable Hardware Chip (EHW) development and many other applications. Electromyography (EMG) signal recorded from healthy and myopathy subjects are non-linear and similar in the time domain as well as in the frequency domain. It becomes very difficult to classify these various statuses. In this paper, we have proposed a feature extraction and classification method of healthy and myopathy EMG signals where two features have been extracted on both healthy and myopathy EMG. Mean Squared Error (MSE) has been calculated in order to observe which feature will give better classification result. Then SVM is used to classify the extracted results. To evaluate the proposed model, a standard dataset collected from physionet.org is used where it shows higher accuracy than the conventional methods.**

***Index Terms*—Electromyography, Muscle Computer Interface, Mean Absolute Deviation, Sample Entropy**

I. Introduction

Electromyography signal is a type of biomedical signal which carries properties of all conventional signals and describable in terms of their amplitude, frequency, and phase. It is originated by the movement of the muscle of the body. The movement of the muscle is known as contraction and vice versa is known as relaxation. It is controlled by the nervous system and depended on the anatomical and physiological properties of the muscle. EMG signals acquire noises while traveling through various tissues and nerves. So recorded EMG signals need to be preprocessed before use. The main purpose of EMG signal classification is in clinical diagnosis and biomedical applications. Neurological and Neuromuscular diseases can also be diagnosed with the help of different EMG signals. In terms of physiological background, there are three types of EMGs- Healthy, Myopathy and Neuropathy or Amyotrophic lateral sclerosis (ALS). Neuropathy or ALS is responsible for neurological diseases and Myopathy is responsible for neuromuscular diseases. Neuropathy or ALS is a rapidly progressive and fatal neurological disease. A statistic based on USA (ALS Association 2016), revealed that approximately 6000 peoples are diagnosed in the United States. So, research on neurological diseases is increasing day by day. Development in this research sector depends on the classification of different EMGs. Better features should be extracted for better classification result. Previous studies related to feature extraction of EMGs have been proposed in four domain- time domain, frequency domain, time-frequency domain and complex network domain. It has been studied that time domain features are easy to extract and work comparatively good in EMG signal classification.

The rest of the paper is organized as follows. Section II presents a brief review of literature related to EMG signal classification and feature selection. Section III presents a detailed description of the proposed approach. A brief description of techniques used in the study and dataset. Section IV presents a description of the experimental set up and evaluation results. Finally, paper ends with the concluding remarks in Section V.

II. Related Works

In 1666, Francesco Redi [1] published documentation where he informed that electricity is generated by the electric ray fish by the highly specialized muscle. In 1773, Walsh demonstrated that muscle tissue of Eel fishes could generate a spark of electricity [2] . In 1792, A. Galvani illustrated that the electricity could commence muscle contraction [3] . Six decades later, Dubios-Raymond found the possibility to record electrical activity throughout a voluntary muscle contraction. Marey made the first recording activity in 1890 and also introduced a new term called electromyography [3] .

Gasser et al. observed the electric signals from muscles using an oscilloscope in 1922 and found only rough information due to the stochastic nature of the myoelectric signal. The detection capability EMG improved in the next 20 years. Researchers used the improved electrodes widely for experiment over muscles. In the early sixties, surface EMG was used for the treatment of particular disorders [1] .

Recently, surface EMG is widely used to examine superficial muscles in clinical protocols. It is used for diagnosis of neurological and neuromuscular problems. It is also used in different research labs such as neuromuscular physiology, biomechanics and so on. When diagnoses are performed based on EMG signals, the system must classify correctly as ALS or myopathy because to handle these two disorders, a number of drugs and therapies are adopted.When studying and developing of this kind, the EMG signal is considered as a unique method to acquire data [3] that records the similar electrical activity of 2 motor units in the neuromuscular operation. EMG signals analysis is usually performed under two cases; for prosthetic device control as well as human-machine interactions [4], [12] and for diagnosing disorders [5] . Neuromuscular disorders can be classified into muscular (myopathy) and neuronal (neuropathy) [6] and they are similar to pathological changes within the structure of the motor unit. The distinguished classification of myopathy and neuropathy is important because it is critical to delimit the treatment due to the differences in the causes of these diseases. To classify problems accurately, a number of feature extraction steps are followed. It is plausible to obtain excellent classification performance if we can extract features sufficiently well.

Initially, the feature extraction of EMG signals was proposed on the frequency domain, the time-frequency domain, and the complex network domain. To extract features in the frequency domain, two spectral model was used - a) Fast Fourier transforms (FFT) and b) Autoregressive (AR) [3] because it can interpret the characteristics of the signal. Nevertheless, each subject carries different strength in their signals. When analyzing the EMG signals, different wavelets are being used for the time-frequency domain [8]–[10] . The ability to analyze in various sub-bands is the main advantages of this method.

However, to analyze the EMG signals, an empirical mode decomposition technique was used for extracting signals. This technique was accepted to generate data in noisy non-stationary and nonlinear processes over an extended quantity of applications [11] .Thus, a visibility algorithm can be used to transform a nonlinear signal into a complex network. A time series signal can be transformed into a graph using a visibility algorithm. The resulting graph bears the characteristics of the time series. For that parpuse, a geometrically insignificant as well as analytically solvable horizontal visibility algorithm was used [12] . Previous works related to complex network analysis are completely theoretical having no indication of implementation in signal analysis. To classify electroencephalogram (EEG) signals, a visibility graph was used from bands with higher frequency to classify electroencephalogram (EEG) signals [5] . Contrasting with the simple entropy method, they concluded that their method was better. In addition, Zhu et al. [5] applied visibility graphs having nonlinear feature extraction algorithms on the EEG signal. But the algorithms are not suitable for practical use because they were slower than FFT analysis. Consequently, Zhu et al. [5] introduced the Fast-weighted Horizontal Visibility Algorithm (FHVA) that can be applied using signals having high amplitude variations. But the classification results by this method are incorrect because of using the horizontal relationship. Due to this reason, the FHVA is not fit for EMG signals.

III. Proposed Work

In the proposed approach, the EMG based classification starts with the preprocessing step. Then the results are calculated using two important steps - Feature extractions and classifications. The detailed description of various modules of the proposed approach is defined in below sections:

A. Data Preprocessing

In order to evaluate the proposed model, dataset first preprocessed in two steps- normalization and framing. In normalization, the signals are normalized with the bandpass filter. The filter has been designed according to needs. Finally, the normalized signals are divided into several frames.

B. Feature Extraction

For each and every frame of signal data, features are extracted and stored in feature vectors. Then, features are compared with each other to determine which feature would give more effective classification result. Finally, a feature matrix is created and classification has been performed over this matrix.

IV. Experimental Setup and Results

This section provides a detailed description of the experiments performed, evaluation metrics, and corresponding results. In addition, it also provides a brief description of the datasets used for the evaluation.

A. Dataset

In order to perform the experimental evaluation of the proposed approach, data has been collected from Physionet for both myopathy and healthy EMG signals. The collected data was normalized data but 6th order Butterworth filter has been designed and imposed according to needs to get the best accuracy. The data were collected by two types of patients-without the history of the neuromuscular disease and with myopathy due to the longstanding history of polymyositis. The data were recorded at 50K Hz and then down sampled to 4 KHz. Throughout the recording process, two analog filters were applied - a 20 Hz high-pass filter and a 5K Hz low-pass filter.

B. Experimental Setup

In order to evaluate the proposed model, we have divided EMG signals into two levels: healthy and myopathy. The dataset is split with a ration of 50% of training and 50% of testing data respectively. Additionally, a fivefold cross-validation is employed in the training and testing data. To evaluate the performance of the classifier, we have to determine the following parameters:

1) Classification result using Mean Absolute Deviation (MAD) only for both subjects
2) Classification result using Sample Entropy (SE) only for both subjects
3) Classification result using both features, which is the total classification on the whole dataset

We observed the performance obtained by the SVM classifier compared with prior works that applied different methods and the total classification accuracy of the proposed method is outstanding.

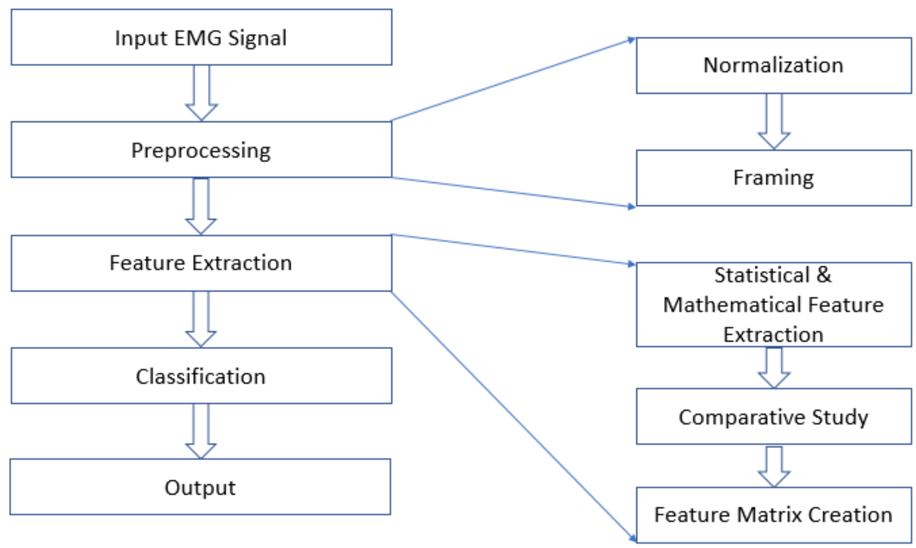

Fig. 1: A schematic representation of the proposed approach for EMG classification

C. Evaluation Metrics

The performance of proposed approach is evaluated using the standard evaluation metrics Mean Absolute Deviation (MAD) and Sample Entropy (SE). The mean absolute deviation of a dataset is the average distance between each data point and the mean. It gives us an idea about the variability in a dataset. One issue of SE calculation is to determine the dimension m and tolerance r.

$$MAD = \frac{\sum |x_i - \bar{x}|}{n} \tag{1}$$

$$SE(x, m, r) = \ln \frac{A^m(r)}{B^m(r)} \tag{2}$$

D. Evaluation Results

In this proposed method, Mean Absolute Deviation (MAD) and Sample Entropy (SE) has been extracted from the collected dataset. Both myopathy and Healthy dataset have been divided into 20 frames, respectively. Each frame contains 630ms data for Healthy EMG signal and 1390ms for myopathy EMG signal. Then each of these two features has been extracted on each frame for both Healthy and Myopathy signals and stored in separate feature vectors respectively.

Figure 4 and figure 5 illustrate the visual differences between the two MAD feature vectors and two SE feature vectors. Mean Squared Error (MSE) also has been implemented to calculate the difference between both feature vectors theoretically. It is observed that MSE has given 0.0023 value for two MAD features and 0.1473 value for two SE features. It is obvious that SE features have more differences then MAD

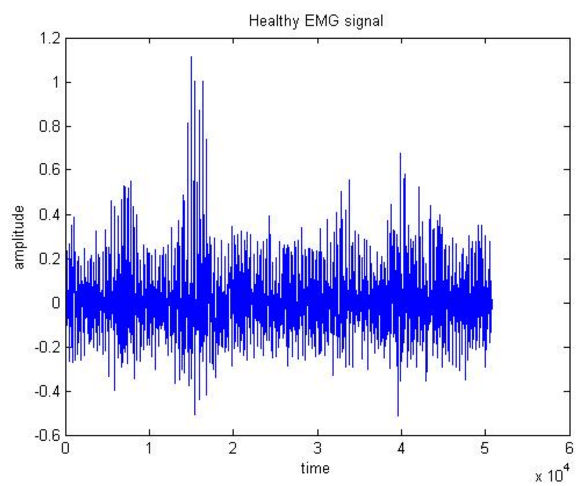

Fig. 2: Healthy EMG plotted data on MATLAB

features. That is why it can be said that SE will provide better classification result than MAD.

E. Discussion

In this paper, we have proposed a method of EMG-based feature extraction using advance effective features for healthy and myopathy detection. Because of the effectiveness of distinct features that are perfectly matched with the patterns of healthy and myopathy signals, our proposed model shows the better accuracy of the classification rather than the conventional method as given in table II. To develop applications in medicine in order to better human life, the outcomes should be

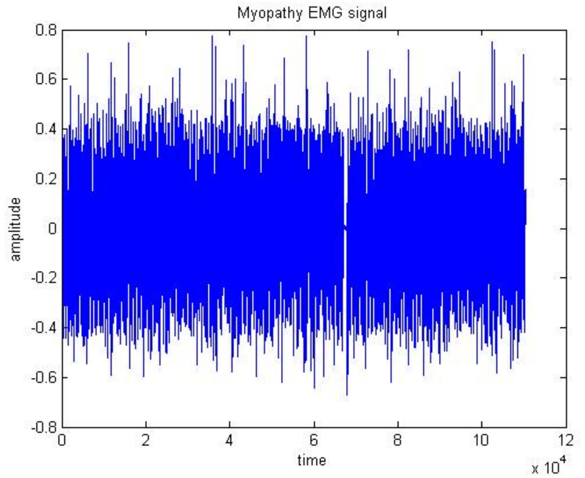

Fig. 3: Myopathy EMG plotted data on MATLAB

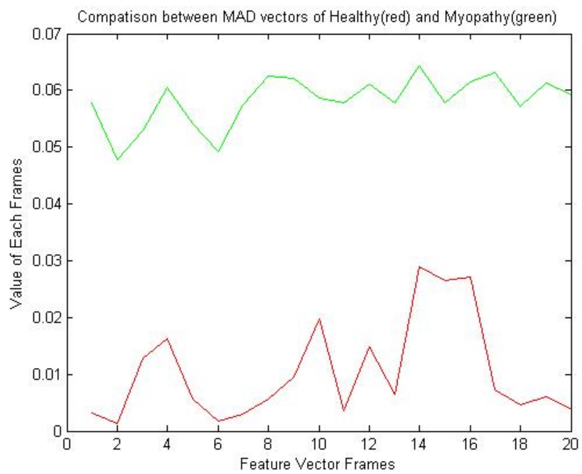

Fig. 4: Comparison between two MAD vectors of healthy (below) and myopathy (above), respectively

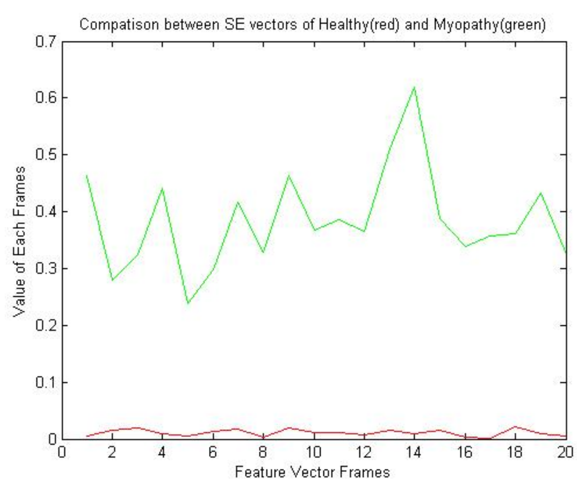

Fig. 5: Comparison between two SE vectors of healthy (above) and myopathy (below), respectively

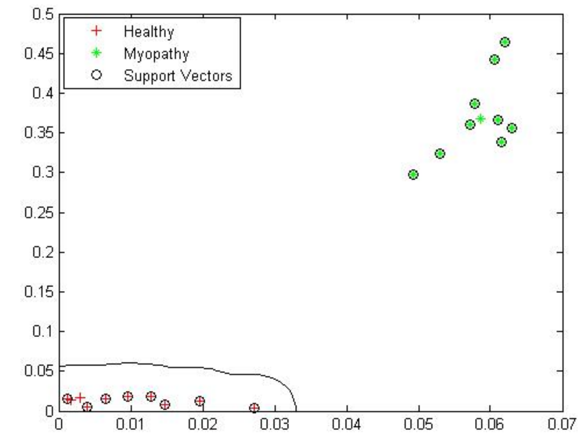

Fig. 6: Training of SVM classifier using non-linear RBF kernel

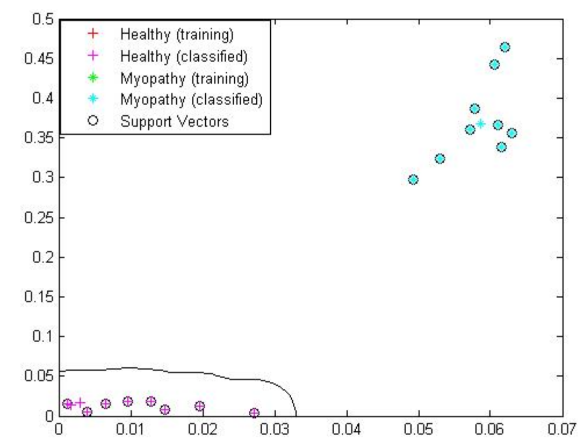

Fig. 7: SVM classification using non-linear RBF kernel

accurate. The result obtained by the proposed method exhibits to the best accuracy that approximated 100%.

In this work, data has been collected from Physionet for both myopathy and healthy EMG signals. The collected data was normalized data but 6th order Butterworth filter has been designed and imposed according to needs to get the best accuracy. Mean Absolute Deviation (MAD) and Sample Entropy (SE) has been extracted from the collected dataset. Both myopathy and Healthy dataset have been divided into 20 frames, respectively. Each frame contains 630ms data for Healthy EMG signal and 1390ms for myopathy EMG signal. Then each of these two features has been extracted on each frame for both Healthy and Myopathy signals and stored in separate feature vectors respectively. Figure 4 and figure 5 show the visual differences between the two MAD feature vectors and two SE feature vectors. Mean Squared Error (MSE) also has been implemented to calculate the difference between both feature vectors theoretically. It is observed that, MSE between has given 0.0023 value for two MAD features

TABLE I: Feature extracted result using MAD and SE for healthy and myopathy

MAD		SE	
Healthy	**Myopathy**	**Healthy**	**Myopathy**
0.0032	0.0577	0.0045	0.4634
0.0012	0.0478	0.0159	0.2785
0.0128	0.0529	0.0189	0.3245
0.0164	0.0605	0.0087	0.4414
0.0055	0.0540	0.0055	0.2388
0.0017	0.0493	0.0136	0.2974
0.0030	0.0573	0.0169	0.4161
0.0057	0.0625	0.0030	0.3291
0.0095	0.0620	0.0189	0.4646
0.0197	0.0586	0.0121	0.3675
0.0035	0.0578	0.0121	0.3861
0.0148	0.0610	0.0080	0.3660
0.0065	0.0578	0.0150	0.5098
0.0289	0.0644	0.0096	0.6183
0.0266	0.0578	0.0143	0.3886
0.0272	0.0616	0.0038	0.3389
0.0073	0.0631	5.2095e-4	0.3565
0.0045	0.0572	0.0208	0.3608
0.0060	0.0613	0.0087	0.4325
0.0039	0.0593	0.0054	0.3234

TABLE II: Performance evaluation results of classification

Subject	MAD	SE	MAD+SE
1	97%	99.98%	99.98%
2	99.98%	99.98%	99.98%

and 0.1473 value for two SE features. It is pretty clear that a Sample Entropy (SE) feature has more differences then Mean Absolute Deviation (MAD) features. That is why it can be said that SE will provide better classification result than MAD.

SVM classifier has been used to classify the data collected from features extraction. 50% of data has been selected for training and 50% for testing using fivefold cross-validation. After this, the classifier has been trained with 50% test data using Non-Linear SVM with RBF kernel. Figure 6 shows the training figure. Non-linear SVM has been used to classify the trained data as figure 7. The classifier showed an unusual result for Sample Entropy and whole dataset but gives a less accuracy for MAD compared with them. Table II shows the classification results.

V. Conclusion and Future Works

EMG signal carries helpful information concerning the nerve system. The aim of this paper is to give brief information about EMG and reveal the different methodologies to distinguish the signal. Feature Extraction and classification techniques for EMG signal are discussed along with their efficiency. This study clearly identifies different types of Feature Extraction techniques of EMG signal so that the proper techniques can be employed during any biomedical research, clinical diagnosis, hardware implementations, and end-user applications. In this paper, we proposed an EMG-based feature extraction model of healthy and myopathy detection where sampling of EMG signals which represent muscle responses was done based on the sampling theory for reversible discrete pulses, and two mathematical models were used for feature extraction from the obtained pulses - Sample Entropy (SE) and Mean Absolute Deviation (MAD). These mathematical features were finally classified using SVM classifiers into normal and myopathic cases. Many different adaptations, tests, and experiments have been left for the future due to the lack of a large dataset. In the future, we can employ this model in three EMGs- Healthy, Myopathy, Neuropathy and finally a system can be made for diagnosis.

References

[1] 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.

[2] F. Sadikoglu, C. Kavalcioglu, and B. Dagman, "Electromyogram (emg) signal detection, classification of emg signals and diagnosis of neuropathy muscle disease," *Procedia Computer Science*, vol. 120, pp. 422–429, 2017.

[3] C. J. De Luca, "The use of surface electromyography in biomechanics," *Journal of applied biomechanics*, vol. 13, no. 2, pp. 135–163, 1997.

[4] H. Manabe, A. Hiraiwa, and T. Sugimura, "Unvoiced speech recognition using emg-mime speech recognition," in *CHI'03 extended abstracts on Human factors in computing systems*. ACM, 2003, pp. 794–795.

[5] X. Zhang and P. Zhou, "Sample entropy analysis of surface emg for improved muscle activity onset detection against spurious background spikes," *Journal of Electromyography and Kinesiology*, vol. 22, no. 6, pp. 901–907, 2012.

[6] M. Nikolic, "Detailed analysis of clinical electromyography signals: Emg decomposition, findings and firing pattern analysis in controls and patients with myopathy and amytrophic lateral sclerosis," Ph.D. dissertation, 2001.

[7] M. Khan, J. Singh, and M. Tiwari, "A multi-classifier approach of emg signal classification for diagnosis of neuromuscular disorders," *J Bioengineer & Biomedical Sci S*, vol. 3, p. 2, 2016.

[8] A. Subasi, "Classification of emg signals using pso optimized svm for diagnosis of neuromuscular disorders," *Computers in biology and medicine*, vol. 43, no. 5, pp. 576–586, 2013.

[9] A. a. Subasi, "A decision support system for diagnosis of neuromuscular disorders using dwt and evolutionary support vector machines," *Signal, Image and Video Processing*, vol. 9, no. 2, pp. 399–408, 2015.

[10] A. Subasi, "Medical decision support system for diagnosis of neuromuscular disorders using dwt and fuzzy support vector machines," *Computers in Biology and Medicine*, vol. 42, no. 8, pp. 806–815, 2012.

[11] P. Artameeyanant, S. Sultornsanee, and K. Chamnongthai, "An emg-based feature extraction method using a normalized weight vertical visibility algorithm for myopathy and neuropathy detection," *SpringerPlus*, vol. 5, no. 1, p. 2101, 2016.

[12] E. Gokgoz and A. Subasi, "Comparison of decision tree algorithms for emg signal classification using dwt," *Biomedical Signal Processing and Control*, vol. 18, pp. 138–144, 2015.