EMG Signal Classification for Detecting Neuromuscular Disorders

Tanvir Ahmed and Md. Kafiul Islam*
Department of Electrical and Electronic Engineering, Independent University, Bangladesh

*Email: kafiul_islam@iub.edu.bd

Abstract. Electromyography (EMG) is a technique for evaluating and recording the electrical activity produced by skeletal muscles to see muscle condition. Nervous system always controls the muscle activities such as relaxation or contraction. An efficient analysis of electromyography (EMG) signals plays an inevitable role in the diagnosis of neuromuscular disorders, prosthesis, and several related applications. Our aim in this study is to differentiate neuromuscular disorder patients from healthy people based on EMG signals. The EMG signals used in this research were recorded from biceps. Artificial Neural Network (ANN) was used for the classification. Eleven features were extracted from the EMG signals for the classification purpose. A comparative analysis was done based on the results. The outcome of this study encourages possible extension of this approach to improve stronger, more resilient and effective implementations.

1. Introduction

1.1. Background and Motivation
EMG is a procedure that evaluates the condition of muscles and the nervous system that controls them [1]. EMG signal is a signal that is complex and it depends on anatomical and physiological muscle attributes [2]. While traveling with different tissues EMG signal makes noise. Moreover, if the EMG detector is located on a particularly deep surface, EMG detector picks up signals from different engine units which can bring about the interaction of the signals. In biomedical engineering, EMG signals became a very remarkable necessity. Until now, many research and efforts have been made in this field, for improvement of algorithms, and the development of old methodologies. Techniques to reduce noise, and the acquisition of correct EMG signals is also developed. EMG signals are widely used for clinical applications as well as prosthetic and rehabilitation purpose. EMG signals are acquired through either surface electrodes [3] or concentric needle electrodes. The former approach is non-invasive, which reveals packages within the exoskeleton and rehabilitation robots, hand movement recognition, and prosthetics. Alternatively, the latter method is hired for diagnosing numerous neuromuscular disorders of neurogenic and myopathy type, which implies that nerves activating a muscle are affected and muscle tissues are affected [9].

EMG signals achieve advanced techniques on the way to find out, decompose, process and classify. Many software programs can gather signals from individual patients, and is able to describe the signal and its nature can be effectively and efficiently [8]. Recent traits in the processing of signal generation and mathematical models have been practiced to improve evaluation and to detect EMG signals. Artificial Intelligence and various mathematical techniques have been involved in this. The parametric
extraction that used from EMG signals is one of the essential steps to find out the feature vector. Feature choice affords the indications for selecting the best feature for classification which depend on several processes. For this reason, feature extraction is very essential for the classification purpose in artificial neural network systems [7]. ANNs have gained a whole lot of achievement over the previous years as an effective approach to resolve many real world issues. The function of ANNs can be used for anticipating and prediction, classification and diagnostic, pattern recognition and estimation and control, system identity, parameter prediction and optimization [7]. A Neural Network (NN) refers to a supervised learning model which means that data needs to be labelled before it is processed. The purpose of a NN is to imitate a biological brain and its immense network of neurons. This is done by using a lot of weighted simply connected nodes. The NN uses these weights in its measurements. NNs also have learning or training algorithms that are used for weight adjustment. ANN has three types of nodes: input, hidden, and output nodes. With each node group organized into a layer where there is no connection between the nodes of the same layer.

We created a neural network to classify healthy subjects from patients with the help of MATLAB. The reason for doing this research is to identify neuromuscular disorders. To identify the patient in conventional way is a lengthy process. If we can build a promisingly accurate neural network, it will be easy to tell if a person has neuromuscular disorders or not. We will only need the MATLAB software installed on our device and we must have the readings taken by an EMG recorder. In this way they can know the test results without any cost. Further study in this topic, will make this network usable in medical purpose, it will make the processing easier, faster and reliable.

1.2. Literature Review
Several works with EMG signals have been performed and the aim is to correctly identify the signals. Classifier is most widely used to distinguish hand gestures for the arm of the prosthesis. Some research work is also being done to classify disorders or to distinguish patients from healthy individuals. It makes clinical applications much easier and faster, only with the assistance of a classifier or network. Authors in [6] analyzed EMG characteristics of the upper limb to diagnose Parkinson's disease where data were collected from 16 PD patients and 25 healthy subjects. In [7], a Multilayer Perceptron (MLP) based ANN is used to diagnose myopathy and neuropathy from healthy EMG signals using 5 features namely Autoregressive (AR), Root Mean Square (RMS), Zero Crossing (ZC), Waveform period (WL) and Mean Absolute Value (MAV). In [8], a comparative study was carried out to diagnose neuropathy using various strategies and algorithms for EMG classification. A single-channel EMG recording is used for diagnosing neuromuscular disorders in [9] where the raw signal is decomposed to various intrinsic modes through EMD followed by applying Fast-ICA before feature extraction. Later a reduced set of five time domain features were extracted from the separated components and then are categorized using the linear discrimination analysis, LDA which received better accuracy. In [10], EMG signals were classified using discriminant analysis and SVM, received excellent average accuracy which is helpful for arm prosthesis research.

Motivated from the above literature review, the objectives of this work are: to find the suitable features to classify EMG signals using a neural network in MATLAB in order to differentiate patients with neuromuscular disorders from healthy subjects. The remainder of this article is organized as follows. The methodology that includes the classes, data acquisition, feature extraction, training, and testing is presented in Section II. The results obtained with a clinical EMG dataset are discussed in Section III. Finally, concluding remarks with potential future works are drawn in Section 4.

2. Materials and Methods
In this study, we have attempted to classify healthy people with people having neuromuscular disorders by creating an ANN network for classification. Figure 1 indicates generic block diagram of this study. System receives two different types of EMG signals. Then, the system decides signal classification to diagnose healthy or patient (neuromuscular disorders). The raw EMG signals were preprocessed before feature extraction.
2.1. Classes

2.1.1. Healthy
A person who do not have damaged muscle or is not affected by a disease, that is, he or she can move freely any time.

2.1.2. Patient
A person having muscular disorders in which the muscle fibres do not function or have damaged nerve or nerve group, resulting in muscular weakness, loss of movement, muscle cramps, stiffness, and spasm. Other symptoms of myopathies can include as muscle cramps, stiffness, and spasm. EMG signals produced by a healthy person are not the same as those generated by a person who has damaged muscle fibre or nerve groups. But in naked eyes it is almost impossible to differentiate the signals as they look very similar.

2.2. Data Collection
For making a classification network or classifier the thing we require after deciding how many class we want, is a dataset. The classifier in this work has been tested with a medical EMG database which is available publicly [5]. The EMG signals were obtained from many subjects with different age. Total of 160 recordings were taken. Unfortunately, we could not work raw EMG signals, the signals were filtered. The high and low pass filters of the EMG amplifier were set at 2 Hz and 10 kHz. There are 7 healthy subjects and 10 patients that are divided into 80 samples as shown in Table 1. Sample raw EMG signals in time domain for both subject types are given in Figure 2.

| Subject type | Muscle type | No. of subjects | No. of samples |
|--------------|-------------|-----------------|----------------|
| Healthy      | Biceps brachii | 7               | 80             |
| Patient      | Biceps brachii | 10              | 80             |

2.3. Feature extraction
Extraction of features is the most important step in creating an ANN network or making any sort of MATLAB classifier. Depending on the signals, we need to extract features. Here, 11 features from each sample were extracted (Total 160 samples). The eleven features are: Root mean square, Waveform length, Variance of EMG, Maximum fractal length, Modified mean absolute Value, Enhanced Wavelength, Difference Absolute Standard Deviation Value, Average Amplitude Change, Variance of FFT, FFT maximum intensity, Variance of neo. Any classifier's accuracy depends mainly on the extracted features. Choosing the features according to the signal is very important.
2.3.1. Root mean square:
The square root of mean value of a squared function is RMS value. RMS is one of the technologies commonly used to evaluate surface EMG signals. It is expressed as:

$$x_{RMS} = \sqrt{\frac{x_1^2 + x_2^2 + \cdots + x_n^2}{n}}$$

(1)

2.3.2. Waveform length:
Given a time series or waveform $v(t)$ with number of $T$ time samples, the change in amplitude between adjacent time samples is known as waveform length. WL is expressed as:

$$WL = \sum_{t=2}^{T} |v(t) - v(t-1)|$$

(2)

2.3.3. Variance of EMG:
The power of an EMG signal is expressed as variance of EMG. It is a useful feature for classification purpose.

2.3.4. Maximum fractal length:
Maximum fractal length is used to characterize fractal patterns.

$$MFL = \log_{10} \left( \sqrt{\frac{\sum_{n=1}^{N-1} (x(n + 1) - x(n))^2}{\sum_{n=1}^{N-1}}} \right)$$

(3)

2.3.5. Modified Mean Absolute Value (MMAV):
Modified Mean Absolute Value is a tool by which muscle contraction rates are detected and measured. It is the moving average of the rectified full-wave EMG signal.

2.3.6. Enhanced Wavelength:
Enhanced wavelength uses multiplexing technique, which multiplexes a number of carrier signals into a single signal by using different wavelengths.

2.3.7. Difference Absolute Standard Deviation Value:
DASDV is a statistic that calculates a data set's dispersion relative to its mean and is measured as the variance's square root. It is determined as the square root of variance by evaluating the difference relative to the mean between each data point [4].

**Figure 2.** A sample raw EMG signal generated from a healthy person (upper trace) and from a patient suffering from Myopathy (lower trace).
\[
DASDV = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1}
\]

2.3.8. **Average Amplitude change:**
Average amplitude is the magnitude average of all instant values in a wave cycle. Average amplitude is the ratio of the sum of all instantaneous values to the number of instantaneous values considered in the phase.

2.3.9. **FFT variance:**
Any physical signal can be broken down into a number of discrete frequencies, or a continuum of frequencies over a continuous range, according to Fourier's analysis. The numerical average of a certain signal or kind of signal is called its spectrum, as evaluated in terms of its frequency material. FFT variance refers to the distribution of spectral energy per unit time.

2.3.10. **FFT maximum density:**
FFT max refers to maximum spectral density.

2.3.11. **NEO variance:**
Non-linear Energy operator algorithm was made to de-noise electrical neuronal signals. The main purpose of NEO is to improve the SNR. If the amplitude of the signal is constant, or zero, the output of the NEO is zero; if the amplitude of the signal is drastically changed and has a high peak, the Neo will go to the highest value.

The NEO applied to a signal \(x[n]\) is given by:
\[
\psi(x[n]) = x[n]^2 - x[n+1]x[n-1]
\]
where, \(n = 1, 2, 3, \ldots\)

2.4. **Training**
Artificial neural network was used for the classification purpose of this project. A neural network is a computational system whose structured architecture, with layers of linked nodes, resembles the networked neuronal structure in the brain. A neural network may learn from data so that patterns can be identified, data categorized and future events predicted can be trained. A neural network divides your input into abstraction layers. It can be trained on many examples, such as the human brain, to recognize patterns in speech or images. The behaviour is characterized by the relation between its individual elements and the power, or weights, of those connections.

In this study, a two-layered feed-forward network with hidden sigmoid and softmax output neurons (patternnet) was used which can arbitrarily classify vectors well, given its hidden layer with enough neurons. The network is trained with a scaled backpropagation of the conjugate gradient (trainscg). We used default settings and parameters. It is possible to set the parameters and settings manually.

In a feed-forward neural network, the extracted features are given as inputs and the outputs are given as binary numbers depending on the classes. The output layer is the same as the number of classes and the number of layers that are hidden depends on the users' choice. From each sample, we extracted 11 features (total 160 samples). An 11x160 array has been fed as the input. The epoch of this process was taken 1000. The number of hidden layer was also taken the same. The output layer was two (no. of classes). As output, an array of 2x160 was fed. Our array was ready to be trained after that. Column 77 to column 84 is shown to clear the idea. Column 77 is the extracted features from the signal of 77 no. sample, whose output is 1 (patient). For column 84, it contains the extracted features from the signal of 84 no. sample, whose output is 0 (healthy).

We need to train the data after feeding the inputs and outputs to the system. To this end, ANN requires three data sets: training, validation, and testing. For the validation and testing scheme, we need to set a percentage. It took 30% (15% for verification and 15% for testing), leaving the remaining 70% for
training. That's how MATLAB trains an artificial neural network. The network itself uses the 15% of the test data to show the network's accuracy.

2.5. Testing
After training our network, we need to test our network for accuracy. We have created an array of 11x100 for this purpose. There were 100 samples in the array (50 patients, 50 healthy) [5]. 11 is the number of features whose sequence is the same as the features extracted from the training section. Number and type of testing samples are given in Table 2.

| Subjects | Muscle type | No. of subjects | No. of samples |
|----------|-------------|-----------------|---------------|
| Healthy  | Biceps brachii | 5               | 09            |
| Patient  | Biceps brachii | 6               | 15            |

3. Results and Discussion
With the help of MATLAB, we have established an artificial neural network that distinguishes patients from healthy subjects. In short, we can tell if a person has neuromuscular disorder or is not using a network. After training the array we created, the network gives a matrix of confusion. This confusion matrix gives us the accuracy of our network. Our network had an overall accuracy of 85%. Training a network in MATLAB has 3 parts: training, validation, testing (This is actually used by the network to see the performance after the training of that network). Here, 80 samples of each group are taken.

3.1. Training Results:
Our network used 52 samples of patients and 60 samples of healthy subjects for the training portion. 45 of the 52 patient samples (7 were identified as healthy) and 54 of the 60 samples (6 were identified as patient) were correctly distinguished in the training of the network. The precision for the identification of the patient and the healthy person in this section is of 88.4 percent.
3.2. Validation Results:
The network used 13 samples of patients and 11 samples of healthy subjects for validation. In the validation of the network, 10 of the 13 patient samples (3 were identified as healthy) and 9 of the 11 samples (2 were identified as healthy) were correctly differentiated. For the identification of the patient and the healthy person in this section is giving an accuracy of 79.2 percent.

3.3. Testing Results:
15 samples from patient and 9 samples from healthy subject for testing part were used by the network. 12 among 15 samples (3 were identified as healthy) for patient and 6 out of 9 samples (3 were identified as healthy) were differentiated correctly in the training of the network. The accuracy for classifying the patient and healthy person in this part is 75 percent.

If all uncertainty matrices are combined, the network can distinguish between 67 samples out of 80 samples (13 were identified as healthy) for the patient and 69 samples out of 80 samples (11 were identified as healthy) for healthy samples. The overall accuracy for the classification of the patient and the healthy person is 85 percent.

After training our network we need to create another array to test it. In our test array, 50 samples were taken from patients and the other 50 was taken from healthy subjects. The size of the array was 11x100, 11 is the no. of features (the same features used for training). 39 out of 50 samples were correctly classified in patient class, which gives an accuracy of 78% while 37 out of 50 samples were correctly classified in patient class, which gives an accuracy of 74%.

So, the average accuracy of healthy and patient subject is 76%. We had a testing accuracy of 76% whereas the testing part of the network had an accuracy of 75%. The error is 1.32%, we can say we got 1.32% higher accuracy when we tested the samples creating an array. But if we compare our result with overall accuracy of 85 percent, the error is 10.6 percent.

A quantitative performance comparison between this work and other works is given in Table 3 and an overall confusion matrix of the classification results is shown in Figure 3.

Table 3. A comparison between recent works and our work

| Ref. Work | Classifier/ Method | Accuracy | Remarks |
|-----------|--------------------|----------|---------|
| [7]       | ANN                | 83.5%    | Different Dataset |
| [8]       | SVM                | N/A      | Different Dataset |
| [9]       | EMD + ICA          | 98%      | Same Dataset, but 3 classes |
| [10]      | SVM                | 98%      | Different Dataset |
| This work | SVM                | 70%      | This dataset, 2 classes |
|           | ANN                | 85%      |         |

3.4. Discussion
In the starting of this work we applied Support Vector Machine classifier to classify patient and healthy subjects. But the accuracy was not that good. After that, we tried to use neural network to classify our subjects, and got an accuracy of 85%, which is acceptable but not useable for medical purposes. Most of the recent works had better accuracy than our classifier, but most of the work was not straight forward. Our goal was to classify patients with neuromuscular disorder and healthy people only by using basic concept of classification, and we successfully achieved that. The network was able to classify a good number of patient and healthy person correctly. We successfully got a basic concept of machine learning
and classification through our work. It will also be easier for others to understand the classification technique, this could be used in future to classify any type of EMG signal.

4. Conclusion
The ANN architecture has been successfully developed for the identification of EMG signals. Neural Network is a useful tool for classifying EMG signals in two different groups, such as healthy, patient signals. These EMG signals produced by muscles in the human body are mostly used for medical field diagnostics. Based on our experimental results, we have found that the correct parameters have been successfully applied to complete the process. The purpose of this work was to present the methodology for the classification of EMG. Getting a drawback or a problem with a system leads to another method developed. This research shows an EMG signal analysis technique for applying the correct method during any biomedical research, clinical diagnosis, applications of end-users and hardware usage. The reported classification accuracy is good. Classification performance is, of course, based on the type of features derived from the EMG recordings.

In order to classify the EMG signals, classification methods such as Support Vector Machine, K-NN, fine-tree was also used. Yet neural network had the best accuracy for our classification among all the classification methods. We had a good accuracy, but the accuracy could be further improved. The accuracy could be slightly increased by applying convolutional neural network. A development of this work will make it useful for clinical applications in the near future, which would be accurate, simple, quick and reliable.

Acknowledgment
The authors are thankful to the Dept. of Electrical and Electronic Engineering (EEE) at Independent University, Bangladesh for the support to perform this work.

References
[1] Basmajian, J.V., 1985. Their function revealed by Electromyography. Muscle Alive, 212.
[2] Kleissen, R.F.M., Buurke, J.H., Harlaar, J. and Zilvold, G., 1998. Electromyography in the biomechanical analysis of human movement and its clinical application. Gait & posture, 8(2), pp.143-158.
[3] Cram, J.R., Kasman, G.S. and Holtz, J., 1998. Chapter 3-Instrumentation. Introduction to surface electromyography, pp.43-79.
[4] Shahid, S., 2004. Higher order statistics techniques applied to EMG signal analysis and characterization (Doctoral dissertation, University of Limerick).
[5] Nikolic, M., 2001. Detailed analysis of clinical electromyography signals: EMG decomposition, findings and firing pattern analysis in controls and patients with myopathy and amyotrophic lateral sclerosis (Doctoral dissertation).
[6] Yang, C.Y., Chen, R.S. and Huang, Y.Z., 2020. Analysis and Discrimination of Surface Electromyographic Features for Parkinson’s Disease during Elbow Flexion Movements. Journal of motor behavior, 52(5), pp.625-633.
[7] Elamvazuthi, I., Duy, N.H.X., Ali, Z., Su, S.W., Khan, M.A. and Parasaruraman, S., 2015. Electromyography (EMG) based classification of neuromuscular disorders using multi-layer perceptron. Procedia Computer Science, 76, pp.223-228.
[8] Sadikoglu, F., Kavalcioğlu, C. and Dagman, B., 2017. Electromyogram (EMG) signal detection, classification of EMG signals and diagnosis of neuropathy muscle disease. Procedia computer science, 120, pp.422-429.
[9] Naik, G.R., Selvan, S.E. and Nguyen, H.T., 2015. Single-channel EMG classification with ensemble-empirical-mode-decomposition-based ICA for diagnosing neuromuscular disorders. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 24(7), pp.734-743.
[10] Alkan, A. and Günay, M., 2012. Identification of EMG signals using discriminant analysis and SVM classifier. Expert systems with Applications, 39(1), pp.44-47.