A New Hybrid Classifier for Neuromuscular Disorders Diagnoses

Nahla F. Abdel Maboud, Bassant Mohamed ELBagoury,
Mohamed Roushdy, and Abdel-Badeeh M. Salem

Computer Science Department, Faculty of Computer & Information Sciences
Ain Shams University, Cairo, Egypt
nahlafarid@yahoo.com, bassantai@yahoo.com, mroushdy@cis.asu.edu.eg, absalem@cis.asu.edu.eg

Abstract

Surface electromyography (SEMG) signal is the recording of electrical activity of a muscle during different actions. Significant features extracted from SEMG provide information for identification of neuromuscular disorders. In this study we proposed a new hybrid classification system based on support vector machine (SVM) and artificial neural network (ANN) in order to classify SEMG signal into normal (NOR), ALS or myopathy (MYO) signal. In order to extract features from the SEMG signal, the signal was analyzed using wavelet transform and then statistical features were calculated and used as input to the classifiers. A comparison was made between SVM classifiers for each group (NOR, MYO, ALS) to select the classifier with higher accuracy. The results showed that SVM classifier for ALS group achieved best accuracy of 98%. Then another comparison was made between ANN classifiers to select the best classifier between NOR and MYO. The results indicated that ANN classifier for MYO group obtained satisfied accuracy of 86.6%. Then the hybrid classifier was constructed with overall accuracy of 85.5%.

Keywords: *Neuromuscular disorders, EMG signal, Hybrid Classifier,Machine Learning, Bio-Medical-Informatics*

1. Introduction

There are two types of muscle diseases : Neuromuscular disorders which include diseases that affect nerves and muscles and causes muscle numbness, weakness and twitching. Amyotrophic lateral sclerosis (ALS) is a motor neuron disease that affects nerve cells in the brain and the spinal cord. And Myopathy (MYO) which includes diseases of skeletal or voluntary muscles and causes theses muscles to become weak but they don't affect the nerve system. Surface electromyography (SEMG) signal shows the electrical activity of a muscle which represents the muscle response during different actions and it provides significant information for identification of various diseases like neuromuscular diseases, muscle degeneration and nerve injury. In order to use SEMG signal as a diagnosis signal, significant features must be extracted from the raw signal as using the raw signal itself in the classification process results in poor classification system with very low accuracy. So that the feature extraction process is considered the most important phase in building a successful diagnostic system. In the last decade, Wavelet Transform (WT) approved its efficiency in analyzing non stationary biomedical signals such as SEMG signals. It transforms the signal into both time and frequency domains. In this study Discrete wavelet transform (DWT) is used for analyzing SEMG signal and extracting significant features which are very useful in identification of healthy, myopathic and neuropathic subjects. In this study five features of

SEMG signal are taken into consideration. Root Mean Square (RMs), Mean Absolute Value (MAV), Zero Crossing (ZC), Slope Sign Change (SSC) and Standard Deviation (SD). These features are then used as inputs for classifiers. The next phase after feature extraction process is the classification phase. Classification is a challenging process. Its challenge is in providing an effective classification technique with better classification accuracy. For this reason various classification techniques have been proposed by many researchers. Many of recent researches used Support Vector machine in SEMG signal classification either for actions and movements identification or for neuromuscular disorders classification. More recent, Gurmanik et al. [3] proposed a technique for diagnosis of neuromuscular disorders based on multi-class support vector machine (SVM) and autoregressive (AR) features. Bassam et al. [4] proposed the use of a committee machines with a Support Vector Machines as the component classifier in order to boost the classification accuracy of multichannel uterine (EMG) signals. Kouta et al. [5] also proposed a classification system for four waist motions by constructing a strong multi-classifier using a combination of four SVMs. SVM was also hybridized with particle swarm optimization (PSO) by Abdulhamit Subasi [2] to improve the EMG signal classification accuracy. There are other classification algorithms have been employed to classify EMG signal such as k-nearest neighbor [6, 7], extreme learning machine (ELM) [8], radial basis function neural networks [9], fuzzy logic and probabilistic neural network [10].

2. Materials and Methodology

2.1 Database Description

The proposed method in this study was tested on a dataset includes real single - channel EMG signals detected from normal, myopathic and neuropathic muscles using a standard concentric needle electrode during low and constant level of contractions [11]. The dataset consists of a normal control group, a group of patients with MYO and a group of patients with ALS. The control group consisted of 5 normal subjects aged 21-37 years, 2 females and 3 males. All of them were in general good shape. None in the control group had signs or history of neuromuscular disorders. The group with MYO consisted of 5 patients; 3 males and 2 females aged 26-63 years. All 5 had clinical and electrophysiological signs of myopathy15. The ALS group consisted of 5 patients, 3 males and 2 females, aged 35-67 years. Besides clinical and electrophysiological signs compatible with ALS, 3 of them died within a few years after onset of the disorder, supporting the diagnosis of ALS. Signals recorded from brachial biceps were selected to test our system as they were the most frequently investigated in the three groups. 15 datasets are utilized from the whole datasets. Each dataset contains a total of 262,134 samples of SEMG signal with a sampling rate of 23,438 samples per second. Thus, the time duration of each of these datasets is 11.184 sec. Each dataset was subdivided into 64 distinct frames, each consisting of 4096 samples.

2.2 Signal Analysis

Wavelet Transform is one of the more efficient techniques for processing non stationary signals such as biomedicalsignals (e.g. SEMG). It transforms the signal into its time-frequency domains. There are two types of wavelet analysis. Discrete Wavelet Transform (DWT) and Continuous Wavelet Transform (CWT). Both of them consume low time for signal processing. CWT is more consistent but DWT approved its efficiency in analyzing non stationary signals although it yields a high-dimensional feature vector. CWT can be expressed by the following equation:

$$\psi(a,b) = \frac{1}{\sqrt{|a|}}\, \psi\left(\frac{t-b}{a}\right) \qquad (1)$$

Parameter a is scale, b is time location and $\psi(t)$ is the 'mother wavelet' which can be taken as a band - pass function.

The factor $\sqrt{(|a|)}$ is ensures energy preservation, which is the same for all values of a and b. The equation of DWT can be given by:

$$x(t) = \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} d(k,l)\, 2^{\frac{k}{2}}\, \psi(2^{-k}t - 1) \qquad (2)$$

where $a = 2k$; and $b = 2kl$; and $d(k,l)$ is a sampling of $W\ (a,b)$ at discrete points k and l.

In our work we applied Daubechies wavelet function of degree four (db4) on each frame of the healthy, MYO and ALS subjects in training and testing data so that the next step is to extract time and time-frequency features from the resulted processed signal.

2.3 Feature Extraction

Successful classification system is mainly dependant on the efficiency of feature extraction stage. There are two approaches for extracting significant information from SEMG signal. Those approaches are spectral and temporal approaches. In this section we will discuss various set of time and time-frequency features we employed in our research.

2.3.1 Mean Absolute Value (MAV)

Mean Absolute Value (MAV) is the average of the absolute value of all time samples. It can be represented by the following function:

$$MAV = \frac{1}{N} \sum_{n=1}^{N} |x_n|, \quad for\ i = 1, \dots, I-1 \qquad (3)$$

The parameter N is the number of samples and I is the number of channels.

2.3.2 Standard Deviation (SD)

Standard Deviation (SD) represents the deviation of the mean value around the origin axis of a given segment of signal. SD can be defined as:

$$SD = \sqrt{E(n-M)^2} \qquad (4)$$

2.3.3 Root Mean Square (RMS)

Root Mean Square is related to standard deviation and is used to calculate constant force and non - fatiguing contraction [1]. It can be defined by:

$$RMS = \sqrt{\frac{1}{N}\sum_{n=1}^{N} x_n^2} \qquad (5)$$

2.3.4 Zero Crossing (ZC)

Zero Crossing (ZC) counts the number of times that the EMG signal changed from positive to negative. A threshold condition should be considered to extract this feature from noisy SEMG signal. This feature provides an approximate estimation of frequency domain properties [1]. It can be expressed by the following function:

$$ZC = \sum_{n=1}^{N-1} [sgn(x_n \times x_{n+1}) \cap |x_n - x_{n+1}| \geq t],$$

$$sgn(x) \begin{cases} 1, & if\ x \geq t \\ 0, & otherwise \end{cases} \qquad (6)$$

t represents the threshold parameter.

2.3.5 Slope Sign Changes

Slope Sign Changes (SSC) is similar to ZC. It represents the number of slope changes between positive and negative among three consecutive segments of the signal [1]. It also requires a threshold condition to avoid the interference in SEMG signal. SSC can be defined by :

$$SSC = \sum_{n=2}^{N-1} \left[f[(x_n - x_{n-1}) \times (x_n - x_{n+1})] \right]$$

$$f(x) \begin{cases} 1, & if\ x \geq t \\ 0, & otherwise \end{cases} \qquad (7)$$

Where t is the threshold parameter.

2.4 Constructing Hybrid Classifier

The proposed classification system was constructed through number of steps. At first three classification modules were considered: ALS identifier, myopathy identifier and normal identifier. Each classifier was fed with the previous extracted features. Then the performance of each classifier was measured and the two classifiers with the higher accuracy were selected for constructing the hybrid classifier. Figure 1 shows the architecture of the proposed system. The most important step in developing a hybrid classifier is to determine types of classifiers to be used. In our system we used two types of classifiers. The first one is the support vector machine (SVM) classifier with Gaussian radial basis kernel function (RBF). SVM is a powerful learning method which aims to find the best the best hyper plane that can separate data perfectly into its two classes.

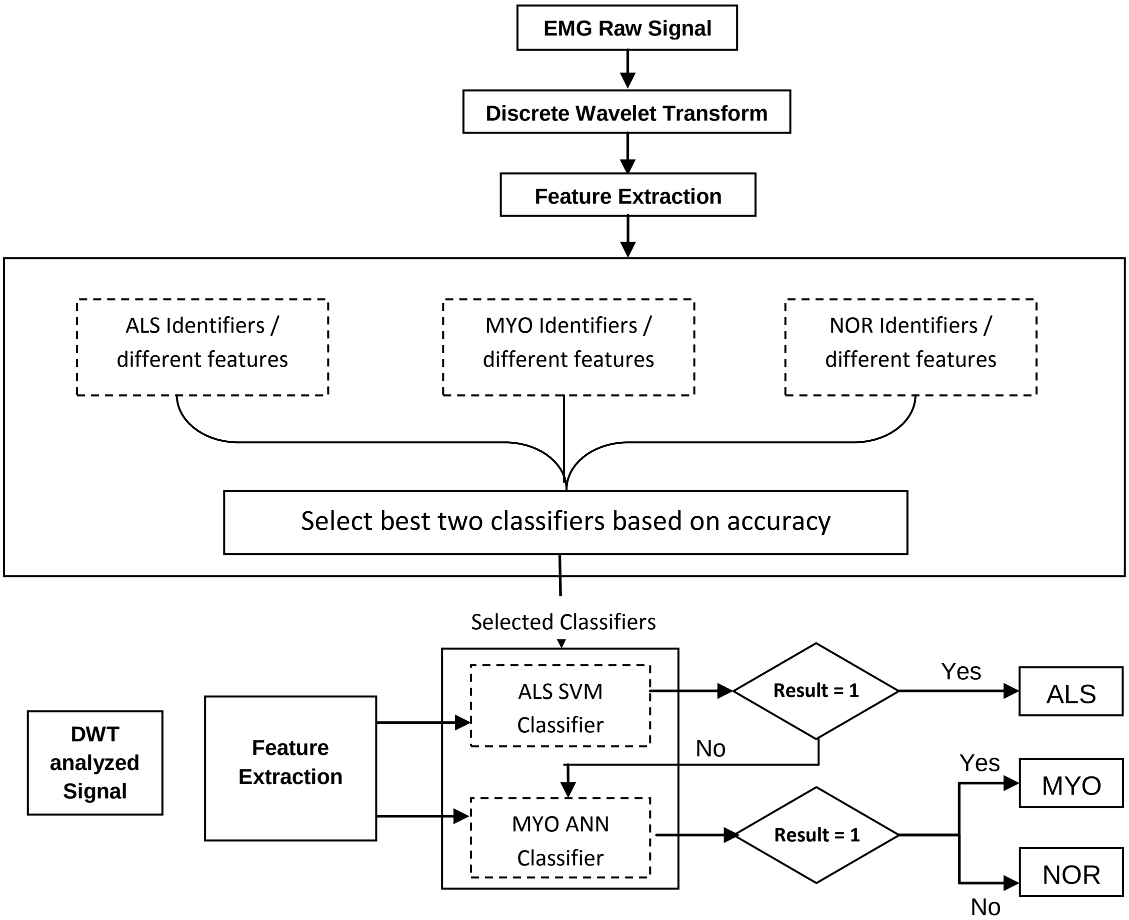

Figure 1. Architecture of the proposed system

Multi classification was recently achieved by combining multiple SVMs. There are two schemes of SVM multi-classifier: a) One Against All which classify each class against remaining classes. b) One Against One which classify between each two classes. RBF used in selected SVM can be represented as

$$K(x_a, x_b) = \exp\left(-\frac{||x_a - x_b||^2}{2\sigma^2}\right) \quad (8)$$

The results shown in table 1 indicate the performance of applying SVM for each group and with different set of features.

Table 1. SVM Classification Table

Feature/Class	ALS	MYO	NOR
MAV	94.9%	66%	67.5%
RMS	98%	65.6%	72%
SSC	67.6%	64%	42%
STD	97.8%	64%	77%
ZC	77%	63%	62.5%
MAV,RMS,SSC,STD,ZC	92.7%	79.6%	78.9%

The second classifier was constructed in order to classifying myopathic subjects against ALS and healthy ones. For this purpose a feed-forward back-propagation based multilayer neural network was used. The designed ANN has 5 neurons in the input layer which represent MAV, RMS, SSC, STD, ZC features. It achieved the best accuracy with 9 tan-sigmoid neurons in the hidden layer and one neuron in the output layer. Selecting number of neurons in the hidden layer is very important issue in building ANN. So that, a comparison was made between the performances of ANN with different number of neurons in the hidden layer. The comparison results are indicated in table 2.

Table 2. ANN Classification Table

Hidden Neurons No./ Class	ALS	MYO	NOR
5	91%	84%	80.8%
6	96%	79.6%	77%
7	96%	84.5%	81.9%
8	93.9%	80%	76.5%
9	95%	86.6%	79%
10	96.6%	81%	78.6%

3. Results and discussion

In our work we made a comparison between SVM classifiers accuracies with each feature to select the classifier with the higher accuracy. As shown in table 1, the higher accuracy for ALS classifiers was obtained by using RMS feature as input to the SVM with accuracy 98%. But MYO and NOR classifiers did not achieve a good accuracy using any of the selected features with SVM classifier. So, a combination of the selected features was used as input to classifiers. A comparison was made between SVM and ANN in classifying MYO using this set of features. As shown in table 1, SVM achieved accuracy of 79.6 % for MYO using RMS, MAV, ZC, SSC and STD features. Another comparison was made for testing ANN with different number of neurons in the hidden layer and selecting the one with the best accuracy then ANN performance was compared with SVM classifier performance. Table 2 shows that the best performance of ANN was achieved by using 9 neurons in the hidden layer with accuracy of 86.6%. As a result, we found that ANN MYO classifier accuracy is higher than SVM classifier. Then the hybrid classifier was constructed using SVM ALS classifier and ANN MYO classifier with overall accuracy of 85.5%.

4. Conclusion

In this study we proposed a new hybrid classifier for identifying healthy, ALS and myopathy subjects based on SEMG signal recordings. Two different types of classifiers with different features were used. First classifier is ALS SVM classifier with RBF kernel function based on RMS feature with accuracy of 98%. Second is MYO ANN classifier with 9 hidden neurons based on RMS, MAV, ZC, SSC and STD features with accuracy of 86.6%. The proposed hybrid classifier achieved overall accuracy of 85.5%.

References

[1]. A. Phinyomark, C. Limsakul, "A Novel Feature Extraction for Robust EMG Pattern Recognition", Journal of Computing, Vol.1, pp 71-80, Dec 2009.

[2]. Abdulhamit Subasi, "Classification of EMG signals using PSO optimized SVM for diagnosis of neuromuscular disorders", Journal of Computers in Biology and Medicine, Vol.43, pp 576-586, 2013

[3]. G. Kaur, A. S. Arora, "Multi-Class Support Vector Machine Classifier in EMG Diagnosis", WSEAS Transactions on Signal Processing, Vol.5, pp 379-389, Dec 2009.

[4]. B. Moslem, M. Khalil, "Combining Multiple Support Vector Machines for Boosting the Classification Accuracy of Uterine EMG Signals", IEEE, pp 631-634, 2011.

[5]. K. Kashiwagi, T. Nakakuki, "Discrimination of Waist Motions Based on Surface EMG for Waist Power Assist Suit Using Support Vector Machine", 50th IEEE Conference on Decision and Control and European Control Conference (CDC-ECC), pp 3204-3209, Dec 2011.

[6]. S. A. Fattah, Md. Asif Iqbal, "Identifying the Motor Neuron Disease in EMG Signal using Time and Frequency Domain Features with Comparison", Signal & Image Processing : An International Journal (SIPIJ), Vol.3, pp 99-114, April 2012.

[7]. S. A. Fattah, Md. Asif Iqbal, "Evaluation of Different Time and Frequency Domain Features of Motor Neuron and Musculoskeletal Diseases", International Journal of Computer Applications, Vol.43, pp 34-40, April 2012.

[8]. Necmettin Sezgin, "Analysis of EMG Signals in Aggressive and Normal Activities by Using Higher-Order Spectra", the Scientific World Journal, Vol.2012, pp 1-5.

[9]. M. O. Diab, B. Moslem, "Classification of Uterine EMG Signals by Using Normalized Wavelet Packet Energy", IEEE, pp 335-338, 2012.

[10]. Shalu George K, K S Sivanandan, "Fuzzy Logic and Probabilistic Neural Network for EMG Classification – A Comparative Study", International Journal of Engineering Research & Technology (IJERT), Vol.1, pp 1-7, July 2012.

[11]. Nikolic M., "Detailed Analysis of Clinical Electromyography Signals EMG Decomposition, Findings and Firing Pattern Analysis in Controls and Patients with Myopathy and Amytrophic Lateral Sclerosis", PhD Thesis, Faculty of Health Science, University of Copenhagen, 2001. [The data are available as dataset N2001 at http://www.emglab.net]