# B. TECH. PROJECT REPORT

**On**

# Automated Classification of Abnormal EMG Signals using Tunable-Q Wavelet Transform


BY

**Abhinav Tripathi**

**Dhaivat Joshi**




**DISCIPLINE OF ELECTRICAL ENGINEERING**

**INDIAN INSTITUTE OF TECHNOLOGY INDORE**


**December 2016**

# Automated Classification of Abnormal EMG Signals using Tunable-Q Wavelet Transform

**A PROJECT REPORT**

*Submitted in partial fulfillment of the requirements for the award of the degrees*

***of***

**BACHELOR OF TECHNOLOGY**
**in**
**ELECTRICAL ENGINEERING**

*Submitted by:*
**Abhinav Tripathi (130002001)**
**Dhaivat Joshi (130002018)**

*Guided by:*
**Dr. Ram Bilas Pachori**
**Associate Professor**
**IIT Indore**


**INDIAN INSTITUTE OF TECHNOLOGY INDORE**
**December 2016**

## CANDIDATE'S DECLARATION

We hereby declare that the project entitled **"Automated Classification of Abnormal EMG Signals using Tunable-Q Wavelet Transform"** submitted in partial fulfillment for the award of the degree of Bachelor of Technology in 'ELECTRICAL ENGINEERING' completed under the supervision of **Dr. Ram Bilas Pachori, Associate Professor, IIT Indore** is an authentic work.

Further, we declare that we have not submitted this work for the award of any other degree elsewhere.

**Abhinav Tripathi** **Dhaivat Joshi**

---

## CERTIFICATE by BTP Guide(s)

It is certified that the above statement made by the students is correct to the best of my knowledge.

**Dr. Ram Bilas Pachori**

**Associate Professor**

**IIT Indore**

# Preface

This report on “Automated Classification of Abnormal EMG Signals using Tunable-Q Wavelet Transform" is prepared under the guidance of Dr. Ram Bilas Pachori.

Through this report we have tried to give a detailed outline of a novel method for the classification of EMG Signals into normal, amyotrophic lateral sclerosis (ALS) and myopathy. This work aims to provide an effective mechanism for identification of neuromuscular disorders at early stage in order to help patients get better treatment and hopefully save many priceless lives. We have reached better accuracy than existing methods.

We have tried to the best of our abilities and knowledge to explain the content in a lucid manner. We have also added graphs and figures to make it more illustrative.

**Abhinav Tripathi**
B.Tech. IV Year
Discipline of Electrical Engineering
IIT Indore

**Dhaivat Joshi**
B.Tech. IV Year
Discipline of Electrical Engineering
IIT Indore

# Acknowledgements

We would like to express our gratitude towards Dr. Ram Bilas Pachori for constantly supporting and motivating us throughout the project and for always giving us new ideas and guiding us in the right direction.

We would like to thank PhD scholar Mr. Rajeev Sharma for all his valuable guidance.

We would also like to thank our families and friends for always being there with us and for being a source of constant motivation.

It is their help and support, due to which we became able to complete the project. Without their support this report would not have been possible.

**Abhinav Tripathi**
B.Tech. IV Year
Discipline of Electrical Engineering
IIT Indore

**Dhaivat Joshi**
B.Tech. IV Year
Discipline of Electrical Engineering
IIT Indore

# Abstract

Neuromuscular diseases can be assessed using Electromyogram (EMG) signals. In this work, a novel technique is presented to classify the EMG signals into various categories such as amyotrophic lateral sclerosis (ALS), normal and myopathy. The proposed method decomposes EMG signals into its constituent motor unit action potentials (MUAP) and then uses different extracted features for training a random forest classifier. Time domain features such as amplitude, rise time are directly applied on the MUAPs while we have reconstructed the signal from constituent sub-band after applying TQWT in order to extract entropy based features. Weka toolbox is used to classify the signals using random forest classifier. Results of two-class classification are shown for ALS versus normal; ALS versus myopathy and normal versus myopathy to establish the effectiveness of the method. The proposed technique provides promising classification accuracy in order to be useful for medical application and help in early detection of the neuromuscular disorders to help many patients find a better way of treatment.

# Contents

## List of Figures



## List of Tables

# Chapter 1

# Introduction

Electromyogram (EMG) is recording of electrical activities of the muscles. The EMG signals are frequently used in human computer interfacing and also have various applications in industrial and clinical fields. These signals exhibit nonstationary nature and includes complex and large variations. EMG signals contain information about the functioning of muscle. It also carry status of the muscle. Such information can be used for the diagnosis of various neuromuscular diseases like: amyotrophic lateral sclerosis (ALS) and myopathy [1].

ALS is a rapid progressive and fatal neuromuscular disease. It severely affects the functioning of both motor neurons. It may lead to degeneration and possible damage to neurons. In this disease, muscles may become smaller and weaker, ultimately resulting in body paralysis. General population affected by ALS is of age group 40 to 70 years [1]. However, it can also affect people of another age groups. Due to ALS, respiratory failure is also possible which may lead to death of the victim, usually after 3-5 years from the onset of disease. Another muscular disease that involves muscular cramp, stiffness, spasm and dysfunction is Myopathy, which affects skeletal muscles' fiber. Myopathy generally stops muscles to work properly. It affects the muscles present at the center of body. However, it does not leads to death of the muscles [1]. Thus, the EMG signals can be used to for the diagnoses and identification of the patients suffering from either of the diseases.

## 1.1 The purpose of this work

There are many neuromuscular disorders that affects the spinal cord, nerves or muscles. From clinical point of view, detection of these diseases at their early stage is an important step for proper cure and treatment of these diseases. Usually, clinical laboratory test is performed for the detecting them. At initial stage, the symptoms of myopathy and ALS appears similar to the symptoms of other diseases. Therefore, it is difficult to diagnosis these diseases at early stages [2]. Recently, EMG signals based techniques are used for the detection of ALS and myopathy diseases. These neuromuscular abnormalities can be identified from the analysis of the recorded EMG signal.

Traditionally, neurophysiologists assess these diseases on the basis of properties of motor unit action potential MUAPs like: shape of the MUAP and its audio characteristics [2]. This shows the importance of

the MUAPs in the detection process of the diseases. Manual detection of the abnormalities using characteristics of MUAPs requires experienced and skilled neurophysiologist. Despite the satisfactory assessment of the MUAPs by the neurophysiologist, the detection of these abnormalities, may not be sufficient for accurate detection of the nuance deviations. Also, the mixed patterns in the MUAPs are also hard to detect with manual assessment. It is, therefore, necessary to perform analysis of MUAPs, quantitatively in order to account for these variability in the abnormal patterns.

The analysis of the time-series exhibiting non-stationary nature can be effectively performed using Wavelet Transform (WT). It provides information of signal in both time and frequency domain, simultaneously [3]-[7]. The proper feature extraction scheme is necessary for the better performance of the classification. Neuromuscular disease classification based on the EMG signal characteristic can be either direct [1] or MUAP based [3]-[7] methods. In direct method, for the purpose of classification, analysis of the EMG signal classification is performed by breaking the signal into non-overlapping frames. The frames can be composed of multiple MUAPs. Now, each frame is used to extract the features and ultimately, the features extracted from each frame are used for the classification. On the other hand, in MUAP based method, firstly, the signal is decomposed into its constituent MUAPs and then, the extracted MUAPs are used for further analysis and feature extraction. The important features of the MUAPs, which carries the information related to neuromuscular disorders are: shapes and firing rates of MUAPs present in EMG signals [8].

## 1.2 Motor unit action potential

The combined unit of muscle fibers and associated motor neurons is known as motor unit [8]. Motor unit generates electrical potential upon electrical or neurological activation. The generation of MUAP is caused by the firing of muscle fibers on the activation of associated motor neuron. Thus, EMG signal is a biomedical signal of complex nature. The complex nature of EMG signal can be attributed to the anatomical and physiological properties of muscles. The nervous system controls the muscles' activity. The EMG signals can be recorded using electrodes placed at skin surface over the muscles or inserting needle electrodes into the muscles. The recording where the electrodes are placed at the skin surface over the muscles can be classified as surface EMG. The recorded EMG signal comprises the combined action potential due to the activities of all the muscle fibers underlying the skin. On the other hand, EMG recorded by inserting needle electrodes is known as intramuscular EMG. Using intramuscular EMG, the action potential generated from individual muscle of the single motor unit can be recorded.

## 1.3 Proposed method

In this work, initially, the EMG signals were decomposed into its constituent MUAPs and then those are used for the classification process. The classification of MUAPs into normal, myopathy and ALS can be performed by analyzing various time based features such as amplitude, duration, area and rise time are used in this work. Also, Tunable Q-factor Wavelet Transform (TQWT), with different Q-values, is applied on the MUAPs, to acquire features corresponding to different frequency components. The features, such as entropies, which measures the complexity of the signals, have been then applied to the different decomposed wavelets. The feature matrix is then given as input to the classifier, which classifies them as Normal, ALS or Myopathy. The block-diagram of the proposed method is depicted in Figure 1.1.

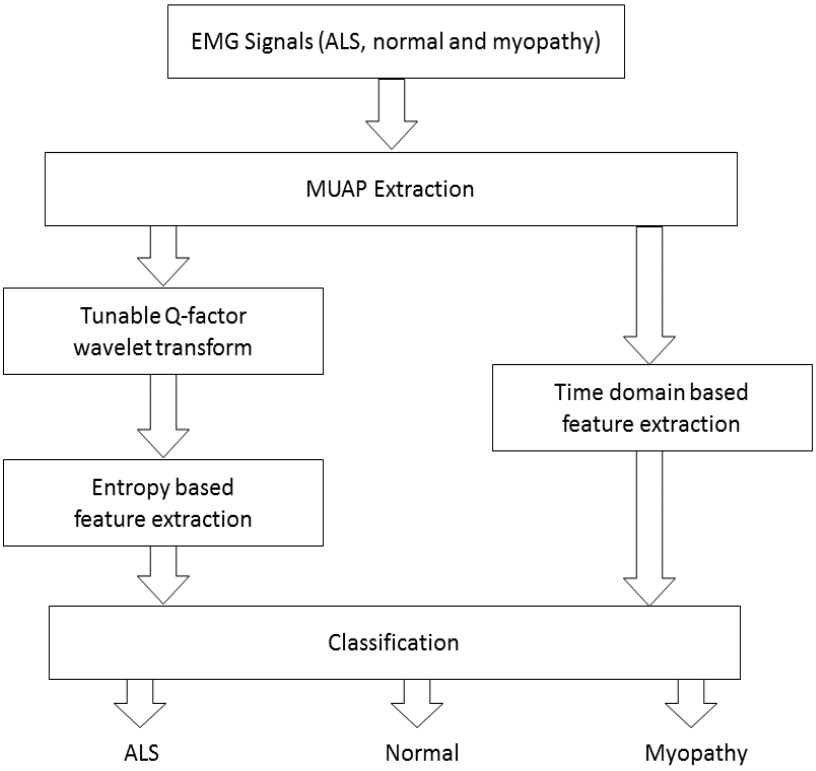


Figure 1.1: Block diagram of the proposed method

# Chapter 2

# EMG Database and MUAP Extraction

The database used in this work and MUAP extraction algorithm from the EMG signal is briefly described in the following section:

## 2.1 Data acquisition and signal filtering

The data was acquired from the EMGLABs [9] database, which is available online. In the database, EMG signals corresponding to three different classes are available. These classes are normal (control), myopathy and ALS. In the control group EMG signals are recorded from 4 females and 6 males subjects with age ranging from 21-37 years. 6 out of 10 were in very good physical shape, and the remaining except one were in general good shape. No subject in the control group had any history of neuromuscular disorders. In the myopathy group, EMG signals are recorded from 2 females and 5 males patients with age ranging from 19-63 years. All 7 patients exhibit sign of clinical and electrophysiological myopathy. In the ALS group EMG signals are recorded from 4 females and 4 males patients with age ranging from 35-67 years. These patients shows the clinical signs compatible with ALS. 5 out of these 8 patients, died within a few years after onset of the disorder [9].

The EMG signals were recorded using concentric needle electrodes. The sampling frequency used was 23,437Hz. The digitization was done using a 16 bit analog to digital (A/D) converter. The amplification factor used was 4000. Filters were used to get the signal in the frequency from 2 Hz to 10 kHz only. Low level of contraction was maintained to acquire to signals. The needles used had a standard leading of area 0.07 $mm^2$.

The typical signals of the three classes are shown in Figure 2.1:

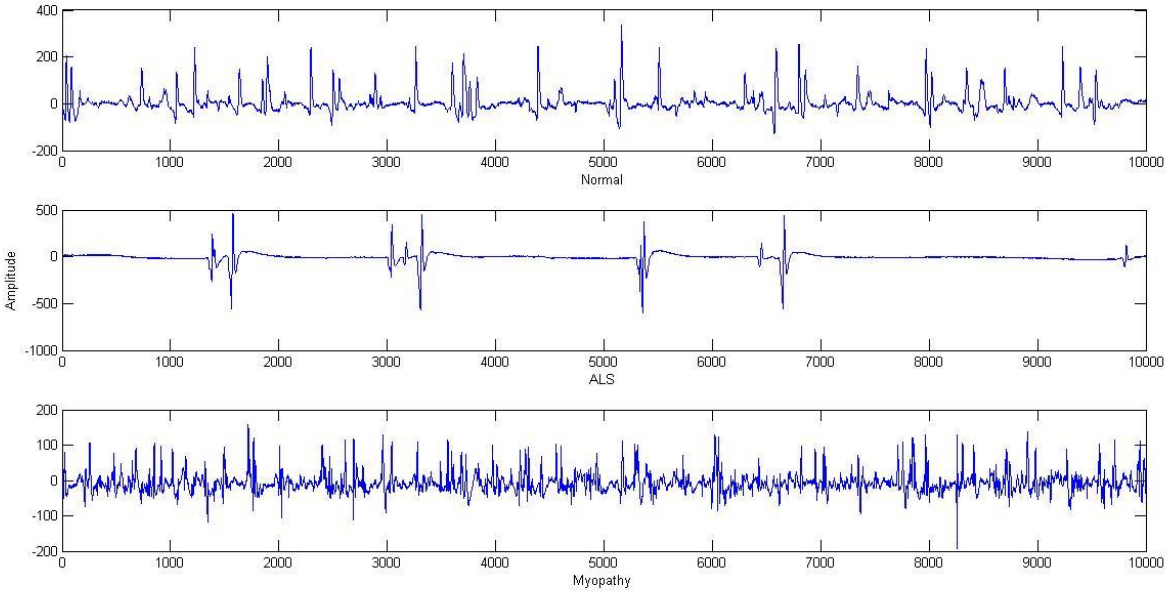


Figure 2.1: EMG signals pattern from top to bottom (a) Normal (b) ALS (c) Myopathy

## 2.2 MUAP extraction

MUAPs are extracted from the EMG signals. The MUAP extraction algorithm consist of two stages namely, segmentation and clustering. The block diagram of MUAP extraction steps is shown in Figure 2.2. The procedure for MUAP extraction is expressed as follows [8]:

1. *Segmentation:* The EMG signal is segmented in the time-intervals containing MUAPs known as segments. A segment can have one MUAP. It can also have superimposed MUAPs known as compound segments. The parts of EMG signal not containing MUAP are known as baseline.
2. *Clustering:* In this step, similar looking segments are grouped together to form a cluster. A group may have multiple number of segments. A group with five or more segments is categorized as potential class (PCL). From each PCL a template is chosen to represent the PCL. A representative template is selected from each PCL. An active MUAP of the EMG signal is represented by this selected template.

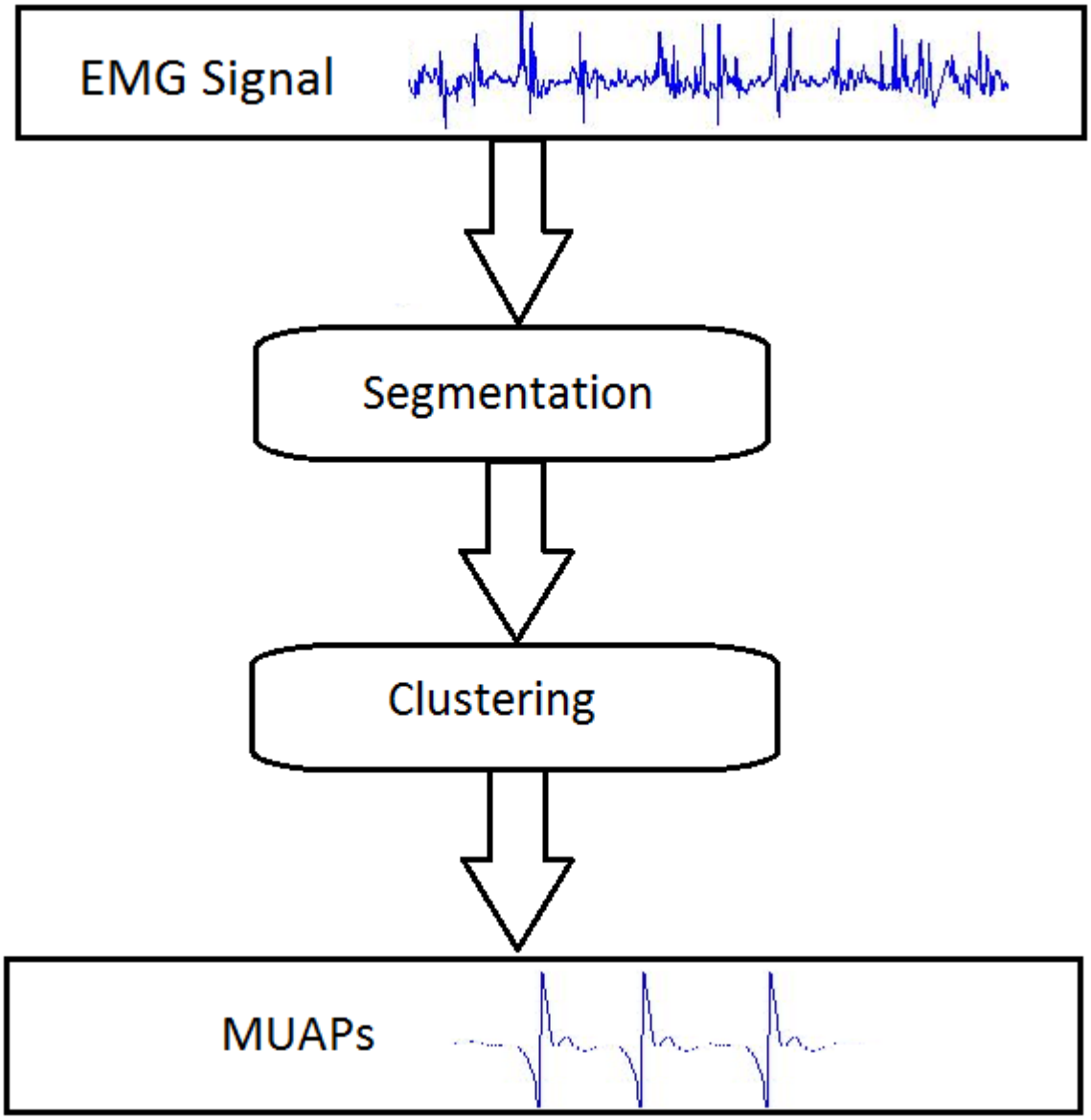


Figure 2.2: Block diagram representation of MUAP extraction: Segmentation and clustering

### 2.2.1 Segmentation

In the segmentation stage, a signal $v(n)$ is partitioned into segments containing MUAP. For the segmentation, a window $N_d$, of 5.6 ms is used and the variance in the window is calculated using [8]:

$$var(j) = \frac{1}{N+1}\sum_{i=-n}^{n} v^2(j+i) - \left(\frac{1}{N-1}\sum_{i=-n}^{n} v(j+i)\right)^2$$

If the variance in a window is more than a threshold *thrd* value, then it is considered as a segment. It may have either a single MUAP or it can be a compound segment consisting of many superimposed MUAPs.

Amplitude density function obtained from the normalized variance signal is used to carry out threshold estimation. A local maximum present in the density function is used to represent similar-sized MUAPs. Threshold is determined as the first local minimum obtained, when starting the search from the origin. It is assumed at the origin, that the smallest MUAPs are assumed to be distinguished from the noise. The segment-delimiting threshold *thrl* is obtained from [8]:

$$thrl = 0.15\ (thrd - blmv) + blmv$$

In the above expression, *blmv* represents the mean obtained from the baseline variance. *blmv* is computed by presegmenting the EMG signal with *thrl* = *thrd*/3.

### 2.2.2 Clustering

The acquired segments are then grouped with similar segments. Similarity between two segments is computed depending upon the distance between two segments. If $s1$ and $s2$ are the two segments, for instance, and $e$ is their difference then the distance between them is given by [8]:

$$dist(s1, s2) = \frac{var(e)}{rms(s1) + rms(s2)}$$

Here, parameter $var$ is the variance of the signal and parameter $rms$ represents root mean square. Before calculating the distance, the signals are time aligned so that we get minimum residual ($e = s1 - s2$). If a group has more than five segments then it can be a potential class. For each potential class, an average segment is chosen which represents the class. These chosen segments represents the MUAPs.

MUAPs have been extracted using the above mentioned method. In the next chapter the signal processing techniques involved to classify the EMG signals into ALS, normal and myopathy groups has been explained.

# Chapter 3

# Tunable-Q Wavelet Transform

TQWT [10] contains a chain of two channel filter banks with low pass output connected to input of the next filter bank. It has the property of perfect reconstruction. It is dependent on three parameters, namely the Q factor, the redundancy and the number of stages used. At every stage, the high pass output is taken as the coefficient for that stage and for the last stage, both the outputs are taken as two separate coefficients.

## 3.1 Parameters of the transform

The parameter $Q$ denotes the amount of sustained oscillations of the wavelet. It is a measure of oscillatory nature of the wavelet. For an oscillatory pulse, we have [10]:

$$Q = \frac{w_c}{BW}$$

where $w_c$ is the centre frequency of the signal and $BW$ is the bandwidth.

The redundancy ($r$) is the over-sampling rate, i.e., the net sampling rate over the input signal's sampling rate. Taking $f_s$ as the sampling rate of the signal, and $\alpha f_s$ and $\beta f_s$ as the sampling rates for low and high pass filters of the two channel filter, then, for a filter of $J$ bands, the filtering rate at sub-band $j$ ($j \geq 1$) is $\beta\alpha^{j-1}f_s$ which when summed over all the levels gives $\frac{\beta}{1+\alpha}f_s$ and hence the over-sampling rate is given by [10]:

$$r = \frac{\beta}{1-\alpha}$$

The number of stages ($J$) denotes the number of filter banks used. Since we have two outputs from the last filter so we have a total of $J$ +1 sub-bands for the wavelet.

An example of 3 stage TQWT filter is as follows:

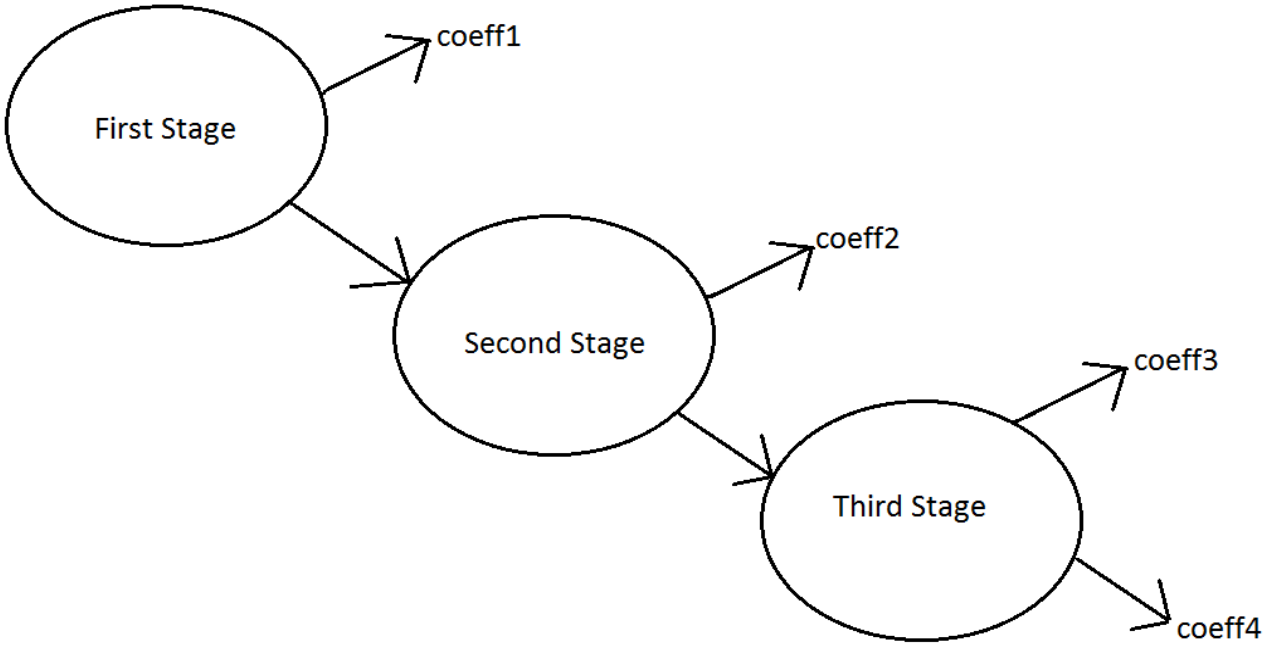


Figure 3.1: Three-staged TQWT using dual channel filter bank [10]

In the Figure 3.1, the high pass output at first and second stage is passed to the next stage and the high pass filter output is taken as coeff1 and coeff2 respectively. Also, both the outputs from the third stage are taken as two different coefficients, coeff3 and coeff4.

## 3.2 Determining the required parameters

We needed to decompose the signal into oscillatory and non-oscillatory components. So, we decided to use TQWT two times with different parameters, once corresponding to the oscillatory component and other time to extract the non-oscillatory components.

The graph of energy of sub-bands for different $Q$ values with the same '$r$' and '$J$' values is shown in Figure 3.2. We found out that for lower $Q$ values, the energy is distributed among all the sub-bands. Also, as we increase the value of $Q$, the energy is concentrated at the sub-band with lower frequencies and the difference between energy levels of low and high frequency sub-bands increases drastically. In the plot shown below in Figure 3.2, sub-band 1 represents highest frequency component and sub-band 11 represents lowest frequency component.

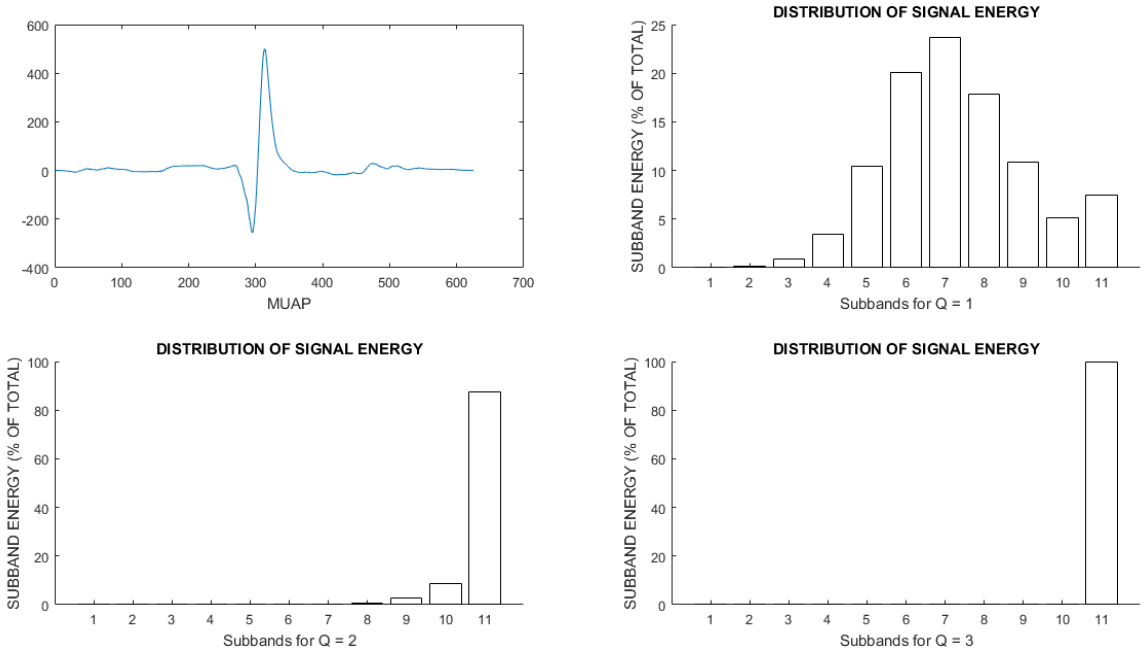


Figure 3.2: MUAP and energy level of sub-bands for different $Q$ values and same $J$ and $r$ values

We decided to use $Q$ = 2 for the second TQWT transform as the energy levels of higher frequency sub-bands were approximately zero for $Q$ > 2.

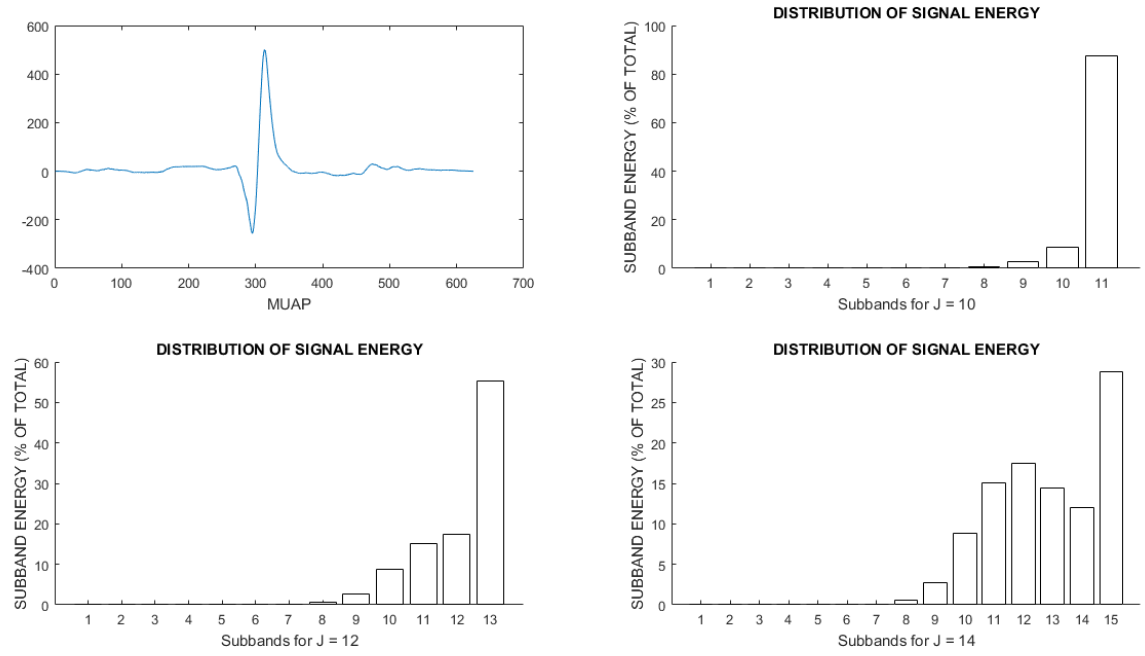


Figure 3.3: MUAP and energy level of sub-bands for different $J$ values and same $Q$ and $r$ values

Also, since the energy distribution depends on the number of levels used in TQWT, so we tried different levels of TQWT and found out that higher levels gave better energy distribution of the sub-bands, as shown in Figure 3.3, was to be used for further processing. The TQWT Toolbox used, however, errors out if the levels used is too high. So an optimal level was found by trial and error.

After looking at the graphs mentioned above Figure 3.2-3.3 and going through many trial and error runs, we decide to use two separate TQWTs, one with $Q = 1; r = 3; J = 10$ and the other with $Q = 2; r = 3; J = 14$. Figure 3.4-3.5 are the signals and their when TQWT with $Q = 1; r = 3; J = 10$ is applied on ALS, normal and myopathy signals:

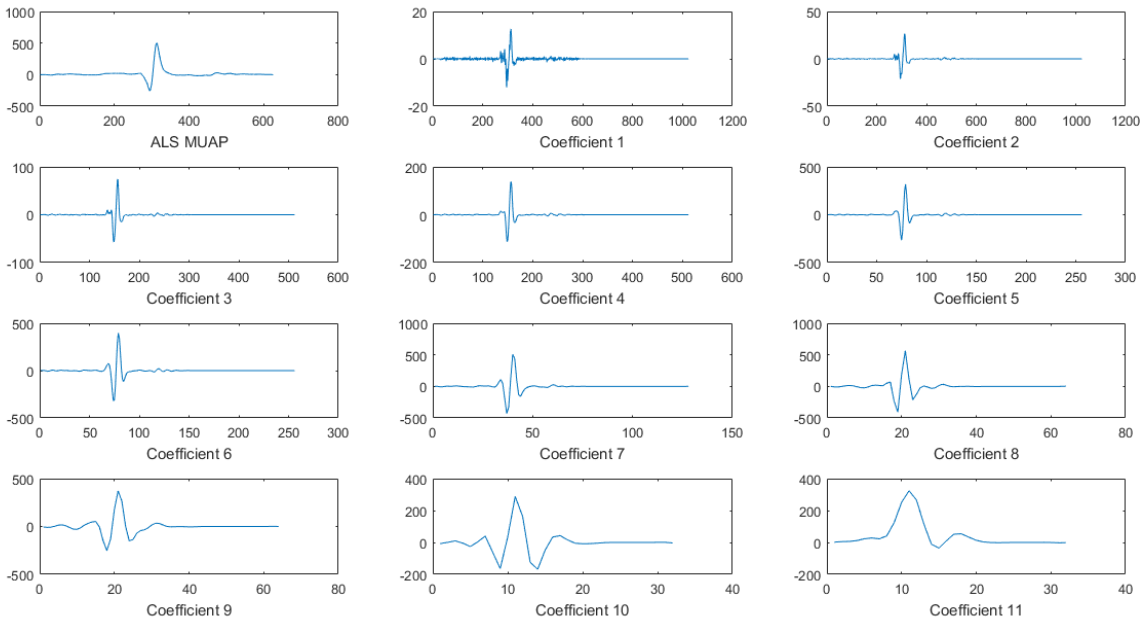


Figure 3.4: Wavelet coefficients obtained after applying TQWT on an ALS MUAP

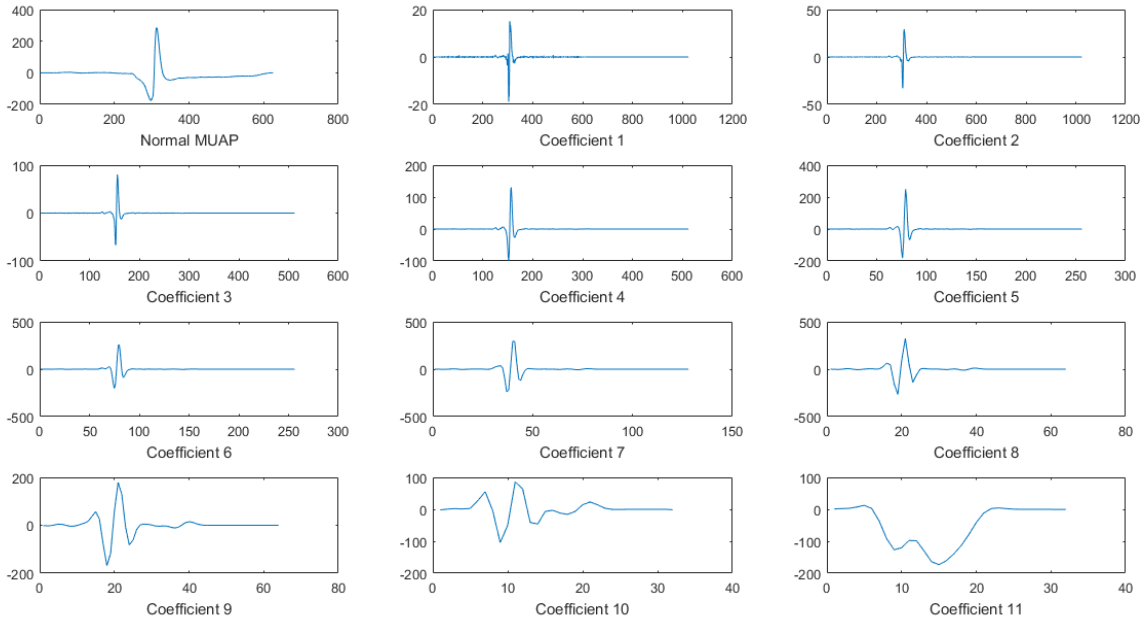


Figure 3.5: Wavelet coefficients obtained after applying TQWT on a normal MUAP

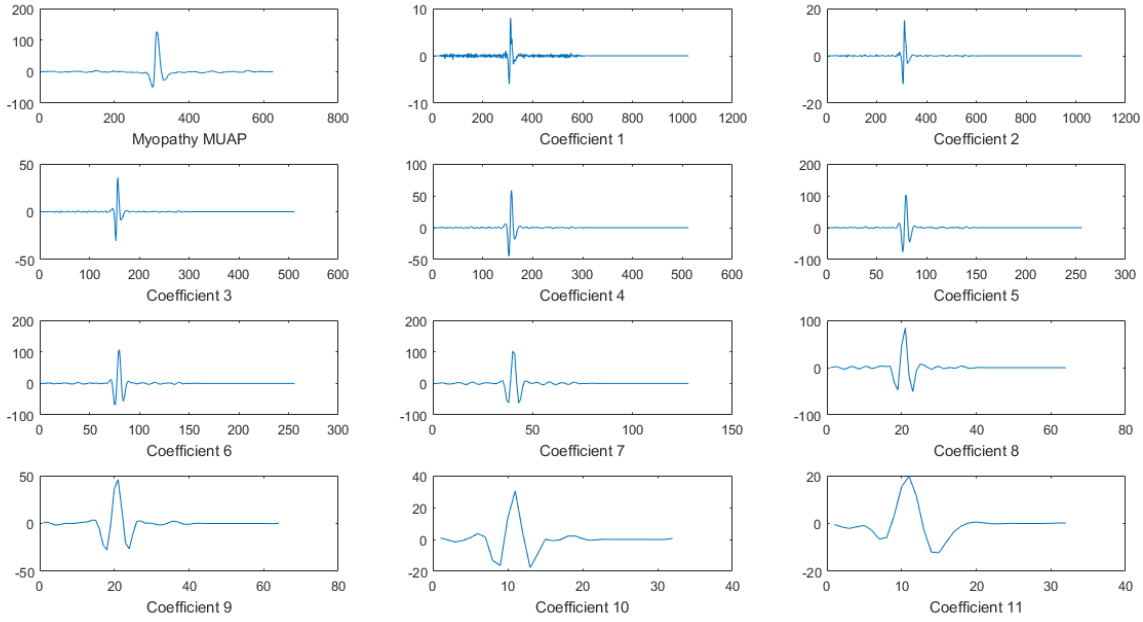


Figure 3.6: Wavelet coefficients obtained after applying TQWT on a myopathy MUAP

# Chapter 4

# Feature Extraction

Feature extraction refers to the procedure of forming a set of measurable properties of the dataset. This helps in reducing the dimensionality of data by extracting a few values which collectively describe the same data, and preferable reduce any unwanted or redundant information. This eases the processing of data. It also helps for better understanding of the data from a human's point of view.

Mostly, when the data to be processed is large, there is a lot of redundant information and leads to waste of processing time. Also, that can lead to confusion during the classification stage. Hence, feature extraction is preferred for classification rather than directly using the input signal values for the same. For any given classification task, we select the desired features such that the desired task can be performed using the most relevant features.

Here, we have used time domain based features alongside entropy based features to cover a large set of detectable features. We have applied several time domain based features directly on the MUAPs and entropy based features on the reconstructed signals obtained from the sub-bands of TQWT.

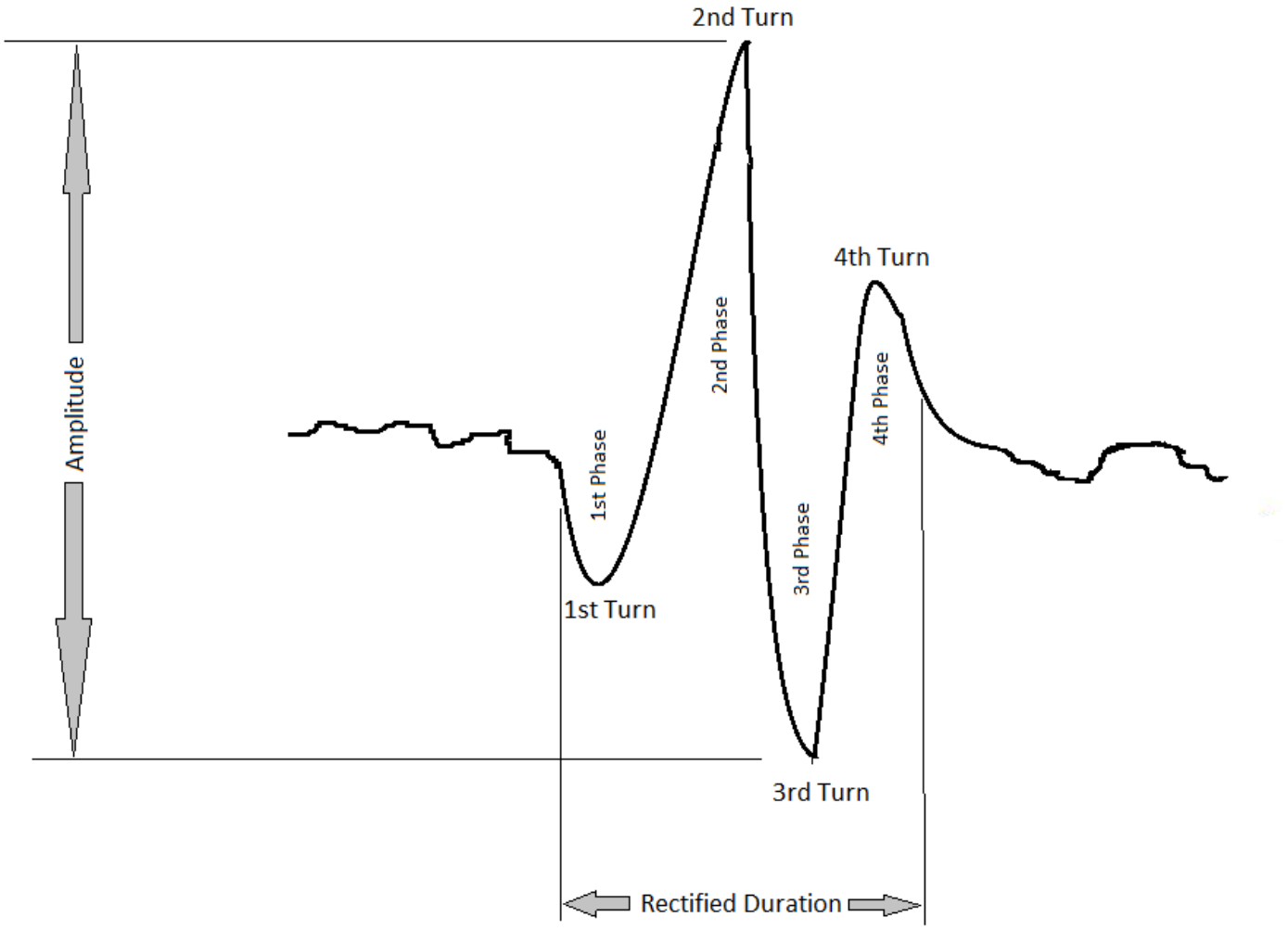


Figure 4.1: Time domain based Features [8]

## 4.1 Time domain based features

The following features, shown in Figure 4.1, are applied on the MUAPs directly:

*i) Amplitude:* Amplitude difference between maximum negative and maximum positive peaks [11]-[12]. The formula is:

$$amp = \max(x) - \min(x)$$

where $x$ is the signal.

*ii) Duration:* To calculate the modified duration of the MUAP from the extracted MUAP, threshold is calculated using the formula [11]-[12]:

$$thres = \frac{\max(x)}{15}$$

The threshold is capped to be between 10 and 20 µV. After this, *(y = thres*) line is drawn and the first and last points where the signal crosses this line are recorded. The time difference between the first and last points where the MUAP crosses this threshold line is defined as the rectified duration.

*iii) Area:* The integral of the MUAP is calculated over the rectified duration obtained [11]-[12]. It is done by using trapezoidal method of finite integration as the MUAP is a discrete signal.

$$area = \int_{t1}^{t2} x(t)dt$$

where *(t1, t2)* is the rectified duration.

*iv) Rise time:* Time difference between maximum positive peak and the minimum negative peak [11]-[12] is used as rise time for this study. It is given as:

$$rt = abs\left(t(x(t) = \max(x)) - \ t(x(t) = \min(x))\right)$$

where *abs* is the absolute value function.

*v) Phases:* We create 2 lines at ± 25 µV and count the number of times the signal crosses these lines, let this be $n_{crossing}$ [11]-[12]. Phases are defined as:

$$phases = \frac{n_{crossing}}{2}$$

*vi) Turns:* Turns are defined as the number of times when the difference between 2 extrema is more than 25 µV. We calculate all the extrema of the signal (both minima and maxima) [11]-[12]. If the difference between 2 consecutive extrema is more than 25 µV then it is counted as a turn.

*vii) Spike duration:* It is the time difference from the first to the last positive spike [12]. It calculates all the times at which positive peaks occur and then calculate the difference between the maximum and minimum values.

$$sp_{dur} = \max(t_{pospeaks}) - \min(t_{pospeaks})$$

where

$$t_{pospeaks} = \left\{ t : x(t) > 0 \; and \frac{dx}{dt} = 0 \right\}$$

*viii) Spike area:* The integral of the MUAP is calculated over the rectified spike duration obtained [12]. It is done by using trapezoidal method of finite integration as the MUAP is a discrete signal.

$$sp_{area} = \int_{t1}^{t2} x(t)dt$$

where *(t1, t2)* is the rectified spike duration.

## 4.2 Entropy based features

Entropy is generally considered as a measure of disorder in a system in thermodynamics. For a given signal, it represents the information content of a signal. In information theory, entropy is used to measure complexity of a time series.

We apply TQWT on MUAPs and then use each of the subbands separately to reconstruct the signal using inverse-TQWT. On each of the reconstructed signal, we apply entropy based features to calculate the complexity inherent in the signal.

The following entropy features [13] are used:

*i) Approximate entropy:* It measures complexity in time domain [14]. It quantifies the probability that a signal will repeat itself. So, with a tolerance *r*, it repeats for *d* points and also for the next *d + 1* points [15]. We define vector X for a given signal x as [13]-[16]:

$$X(i) = \{ x(i), x(i+1), \ldots, x(i+d+1) \} \;\; 1 \le i \le N - d + 1$$

Approximate entropy is then defined as follows [16, 17]:

$$ApEn(d, N, r) = \varphi^{d}(r) - \varphi^{d+1}(r)$$

where $\varphi$ is defined as [13]-[16]:

$$\varphi^d(r) = \frac{1}{N-d+1}\sum_i C_i^d(r)$$

where *C* is the correlation integral for two vectors *X(i)* and *X(j)* which is defined as [13]-[16]:

$$C_i^d(r) = \frac{1}{N-d+1}N_i^r \qquad i = 1, 2, 3, \ldots, N-d+1$$

where $N_i^r$ represents the number of vectors *X(j)* whose distance from the vector *X(i)* is less than *r*.

*ii) Sample entropy:* Sample entropy builds upon *ApEn* to measure complexity of a series. It is independent of length of the signal and helps improve the consistency by removing the self matches than occur in calculation of *ApEn*. Using the vectors *X(j)* as defined for *ApEn*, the sample entropy is defined as [13]-[16]:

$$SpEn(d, r, N) = -\ln\left(\frac{A^d(r)}{B^d(r)}\right)$$

Where:

$$A^d(r) = \frac{1}{N-d}\sum_{i=1}^{N-d} C_i^{d+1}(r)$$

$$B^d(r) = \frac{1}{N-d}\sum_{i=1}^{N-d} C_i^{d}(r)$$

and

$$C_i^d(r) = \frac{1}{N-d}\,C_i \qquad i = 1, 2, \ldots, N-d$$

where $C_i$ is the count such that distance between *X(i)* and *X(j)* does not exceed *r*.

*iii) Spectral entropy:* Spectral entropy is evaluated using normalized Shannon entropy and measure complexity of the signal in frequency domain. It uses Fourier transform to get the power spectral density of the signal in order to calculate energy and hence entropy. Normalized power is calculated by dividing power of each component by total power as [13]:

$$p_f = \frac{P_f}{\sum P_f}$$

Then spectral entropy is defined as [13]:

$$ShEn = \sum_{f} p_f \log\left(\frac{1}{p_f}\right)$$

*iv) Renyi entropy:* It measures spectral complexity of the time series. It is measured by using normalised Shannon entropy. The formula for Renyi's entropy is [13]:

$$RenEn(\alpha) = \frac{1}{1-\alpha} \log\left(\sum_{i} p_f^{\alpha}\right)$$

In this work, we use Renyi's quadratic entropy, i.e., *RenEn* with $\alpha = 2$,

$$RenEn(2) = -\log\left(\sum_{i} p_f^{\alpha}\right)$$

*v) Fuzzy entropy:* It measure the fuzziness of a data by using the distance measure. It measures the similarity in terms of subset-hood (sub-message-hood) of a data. It is also, sometimes, used as conditional probability.

We define vector *X* for a given signal *x* as [18]:

$$X(i) = \{\, x(i), x(i+1), \ldots, x(i+d+1) \,\} - x0(i) \;\; ; \;\; 1 \le i \le N-d+1$$

Where x0 is defined as [18]:

$$x0(i) = \frac{1}{m} \sum_{j=0}^{d-1} x(i+j)$$

Fuzzy Entropy is then defined as follows [18]:

$$FzEn(d, N, r) = \ln(\varphi^d(n, r)) - \ln(\varphi^{d+1}(n, r))$$

where $\varphi$ is defined as [18]:

$$\varphi^d(r) = \frac{1}{N-d} \sum_{i=1}^{N-d} \left( \frac{1}{N-d-1} \sum_{j=1, j\neq i}^{N-d} D_{ij} \right)$$

where $D_{ij}$ is the exponential [18]:

$$D_{ij} = \exp\left\{-\frac{(L_{ij})^n}{r}\right\}$$

Here, $L_{ij}$ is the distance between vector *X(i)* and *X(j).*

The feature matrix hence obtained is then sent to the classifier to classify the data into the required categories. It contains a total of 138 features, 8 time-domain features and 5 entropy features on each of the 26 reconstructed signals from the two TQWT wavelets.

# Chapter 5

# Classification

Classification is used to identify the group in which a new observation belongs. This identification is based upon apriori knowledge of a dataset with similar observations for which we know the category of each observation. For example, a dataset of cancer patients containing information about doses, recovery etc for patients can be used to identify if a new patient will make recovery given the doses of drugs that he is subjected to.

It is considered as supervised learning in the field of machine learning [19] because we use a training dataset for which we already know the correct results. If it is unsupervised then it is called clustering which involves finding out inherent similarity or distance in the data for grouping into various categories.

Features, described as a set of quantifiable properties, which may be integral or real valued, are used to classify the set of observations into different groups. Some classifiers use a distance function to compare an observation to previously known observations.

A classification algorithm is known as a classifier. It is basically a mathematical function that finds the most likely group to which an observation belongs. The classifier that we are using is called Random Forest Classifier.

## 5.1 Random forest classifier

The method of random decision forests [20] uses random decision trees, which constitute of a forest, and restricts them to fixed feature dimensions. Upon such restriction, this method obtains higher accuracy and also doesn't suffer from overfitting. It has also been shown that any method with splitting trees gains the same benefits if it randomly forces the trees to be insensitive to some features [21]. An explanation for random forest classification to be resistant to overtraining can be given using Kleinberg's theory of stochastic discrimination.

It incorporates the idea that for continuing the growth of a decision tree, randomly visiting the available decisions leads to more natural decision making process. This is further extended by using a set of tress, among which, the training data is randomly distributed to fit for each node/tree. Since this method uses a set of trees for the classification procedure, hence it is called 'random forest classifier'.

The use of many random decision trees is quite beneficial as even if a test data is wrongly classified by a few of the trees, taking the mode of the results by all the trees helps in avoiding false results.

Due to all these benefits, we chose to use this classifier for our dataset.

## 5.2 Weka classification toolbox

Waikato environment for knowledge analysis (Weka) [22] is a machine learning software developed at the University of Waikato, New Zealand. It is widely used for educational and research purposes as it is a free software and provides a wide variety of machine learning and data mining tools. It has a lot of features and can be used to perform classification, among many other things, given a set of features.

It was aimed towards agricultural domains in its earlier days but later on it started including tools that have application in many different areas. Moreover, its portable nature and helpful GUI has contributed in its wide acceptance.

A feature matrix was obtained after applying all the features on the MUAPs. It was written in arff format which is understood by weka. Weka toolbox was later used to classify the dataset using random forest classifier.

We used 66% of the data to train the classifier and rest of the data was used for testing. 10-fold verification was also used to verify the results.

# Chapter 6

# Results and Discussion

In this chapter, we have included the results obtained by applying proposed method on the EMGLabs database. The results are presented and briefly discussed in the following sections.

## 6.1 Results

In the proposed method, firstly, MUAPs are extracted from the EMG signals. From the database, a total of 4112 MUAPs are extracted from total of 935 EMG signals. 1051 MUAPs are extracted from 320 EMG ALS signals; 1429 MUAPs are extracted from 300 normal EMG signals; and 1632 MUAPs are extracted from 315 myopathy EMG signals. The typical examples of MUAPs extracted from the EMG signals of normal, ALS and myopathy class are shown in Figure 6.1. The MUAP segments are decomposed using TQWT. The TQWT parameters used to decompose EMG signals are: $Q$ = 1; $r$ = 3; $J$ = 10 and $Q$ = 2; $r$ = 3; $J$ = 14. The entropy features are extracted from the obtained sub-bands. Several time based features like amplitude, time duration, area, turns, phases, spike duration and spike area are computed directly from MUAPs without TQWT decomposition. The final features set is formed by combining all the features computed from the TQWT sub-bands and time based features computed from MUAPs. The size of input feature set is 4112x138. For the purpose of classification, random forest classifier has been used. For the performance evaluation, total number of MUAPs are divided in two parts: 66% for training and 34% for testing. In other case 10-fold cross-validation is also performed. In machine learning, a classification is evaluated on the basis of many factors. Accuracy is the most important factor for measuring the success of the classification. Accuracy is defined as the ratio of number of correctly classified subjects to number of total subjects. For medical purposes, specificity is considered to be the dominant factor in determining the usefulness of a given method. Specificity represents the number of correctly classified normal subjects to the total number of normal subjects. Healthy person is expected to not be falsely told to have disease and waste his time and resources on treatment of a disease which he doesn't even affected by. The results obtained for 66% training data and 34% testing data are presented in Table 6.1. The results obtained for 10 fold cross-validation are given in Table 6.2.

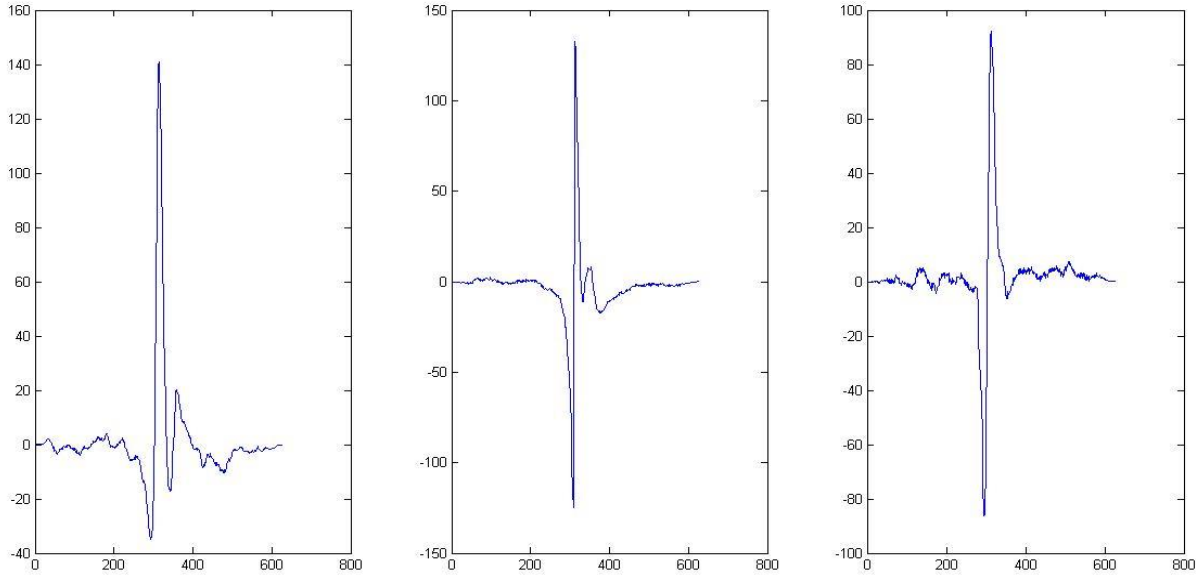

Figure 6.1: Sample MUAPs extracted from EMG signal from left (a) ALS (b) Normal (c) Myopathy

Table 6.1: Classification accuracy and specificity of different class group using 66% training and 34% testing data

| | **ALS versus Myopathy** | **ALS versus Normal** | **Myopathy versus Normal** |
|---|---|---|---|
| Accuracy | 93.58% | 89.16% | 82.41% |
| Specificity | - | 95.65% | 76.60% |

Table 6.2: Classification accuracy and specificity of different class group using 10 fold cross-validation

| | **ALS versus Myopathy** | **ALS versus Normal** | **Myopathy versus Normal** |
|---|---|---|---|
| Accuracy | 94.42% | 87.40% | 81.14% |
| Specificity | - | 94.92% | 81.77% |

## 6.2 Discussion

We have considered three cases of two class classification. In first case classification is performed between ALS and myopathy. In the second case, classification is performed between ALS and normal and in the third case, classification is performed between myopathy and normal. The classification accuracy for ALS versus myopathy was highest. The obtained classification accuracy is significant improvement over previous work [1].

This can enable detection of exact disease for a patient whose muscles are not responding well and hence can help decide the path for further treatment. This will enable save lives for myopathy patients. For ALS patients, this can help decide the right way to proceed as ALS is lethal and patients consider different factors such as monetary factors to decide the type of treatment to take. The accuracy for normal versus ALS and normal versus myopathy is also good and can help doctors to decide if the patient is healthy or not. In case the patient is not healthy then the correct type of disease can be found using ALS versus myopathy classifier.

# Chapter 7

# Conclusion and Future Work

In this work, a method is presented to detect ALS and myopathy neurological disorders from EMG signals. TQWT is used for the decomposition of the MUAPs. The entropy features extracted from the sub-bands of MUAPs are found significantly effective for the detection and diagnosis of the neuromuscular diseases. The results obtained with the proposed methodology is compared with existing method. The results shows the superiority of the proposed method over existing method.

This work improves upon existing literature for particular disease detection. The accuracy for normal versus abnormal cases can be improved in further works. This work still provides significant results to be useful for practical purposes. Once a patient is confirmed to have deteriorating muscles, then too the confirmation regarding the type of disorder can help a lot to make proper choice for getting better treatment and out work provides significant result for classification of the 2 types of disorders (ALS and myopathy).

Future work comprises of, making a system that could allow real-time acquisition and classification of signals in order to help doctors diagnose patients, make a suitable GUI for the system to allow ease of access, exploring different transforms which might help in bringing out better features to work with, and exploring other frequency domain features which might help in better classification of the signals.

# References

[1] V. K. Mishra, V. Bajaj, A. Kumar, Classification of normal, ALS, and myopathy EMG signals using ELM classifier, AEEICB Conference (2016).

[2] A. Fuglsang-Frederiksen, The utility of interference pattern analysis, Muscle and Nerve, 23 (1) (2000) 18-36.

[3] F. G. Nihal, K. Sabri, Classification of EMG signals using PCA and FFT, Journal of Medical Systems, 29 (3) (2005) 15-25.

[4] E. W. Abel, H. Meng, A. Forster, D. Holder, Singularity characteristics of needle EMG IP signals, IEEE Transactions on Biomedical Engineering, 53 (2) (2006) 219-225.

[5] M. R. Canal, Comparison of wavelet and short time Fourier transform methods in the analysis of EMG signals, Journal of Medical Systems Springer, 34 (1) (2010) 91-94.

[6] K. Englehart, B. Hudgin, P. A. Parker, A wavelet based continuous classification scheme for multifunction myoelectric control, IEEE Transactions on Biomedical Engineering, 48 (3) (2001) 302-311.

[7] T. Y. Fukuda, J. O. Echeimberg, J. E. Pompeu, P. R. G. Lucareli, S. Garbelotti, R. O. Gimenes, A. Apolinario, Root mean square value of the electromyographic signal in the isometric torque of the quadriceps, hamstrings and brachial biceps muscles in female subjects, Journal of Applied Research, 10 (1) (2010) 32-39.

[8] M. Nikolic, C. Krarup, EMGTools, an adaptive and versatile tool for detailed EMG analysis, IEEE Transactions on Biomedical Engineering, 58 (10) (2011) 2707-2718.

[9] K. C. McGill, Z. C. Lateva, H. R. Marateb, EMGLAB: an interactive EMG decomposition program, Journal of Neuroscience Methods, 149 (2) (2005) 121-133.

[10] I. W. Selesnick, Wavelet transform with tunable Q-factor, IEEE Transactions on Signal Processing, 59 (8) (2011) 1-18.

[11] C. I. Christodoulou, C. S. Pattichis, Unsupervided pattern recognition for the classication of EMG signals, IEEE Transactions on Biomedical Engineering, 46 (2) (1999) 169-178.

[12] A. Goen, Classification of EMG signals for assessment of neuromuscular disorders, International Journal of Electronics and Electrical Engineering, 2 (3) (2014) 242-248.

[13] N. Kannathal, M. Choo, U. R. Acharya, P. Sadasivan, Entropies for detection of epilepsy in EEG, Computer Methods and Programs in Biomedicine, 80 (2005) 187-194.

[14] S. Pincus, Approximate entropy as a measure of system complexity, Proceedings of the National Academy of Sciences of the USA, 88 (1991) 2297-2301.

[15] D. Lake, J. Richman, M. Grison, J. Moorman, Sample entropy analysis of neonatal heart rate variability, American Journal Physiology-Regulatory, Integrative and Comparative Physiology, 283 (2002) 789-797.

[16] J. Richman, J. Moorman, Physiological time-series analysis using approximate entropy and sample entropy, American Journal of Physiology – Heart and Circulatory Physiology, 278 (2000) 2039-2049.

[17] D. Cui, J.Wang, Z. Bian, Q. Li, L.Wang, X. Li, Analysis of entropies based on empirical mode decomposition in amnesic mild cognitive impairment of diabetes mellitus, Journal of Innovative Optical Health Sciences, 8 (2015) 1-20.

[18] W. Chen, Z. Wang, H. Xie, W. Yu, Characterization of surface EMG signal based on fuzzy entropy, IEEE Transactions on Neural Systems and Rehabilitation Engineering, 15 (2) (2007) 266-272.

[19] Alpaydin, Ethem, Introduction to Machine Learning, MIT Press, (2010) 9.

[20] T. K. Ho, The random subspace method for constructing decision forests, IEEE Transactions on Pattern Analysis and Machine Intelligence, 20 (8) (1998) 832–844.

[21] L. Breiman, Random Forests, Machine Learning, 45 (1) (2001) 5–32.

[22] E. Frank, M. A. Hall, and I. H. Witten, The Weka Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques", Morgan Kaufmann, Fourth Edition, 2016.