A FEATURE EXTRACTION SCHEME BASED ON SPECTRO-TEMPORAL ANALYSIS OF MOTOR UNIT ACTION POTENTIAL OF EMG SIGNAL FOR NEUROMUSCULAR DISEASE CLASSIFICATION

by

Abul Barkat Mollah Sayeed Ud Doula

MASTER OF SCIENCE IN ELECTRICAL AND ELECTRONIC ENGINEERING

Department of Electrical and Electronic Engineering
BANGLADESH UNIVERSITY OF ENGINEERING AND TECHNOLOGY
August 2013
The thesis entitled “A FEATURE EXTRACTION SCHEME BASED ON SPECTRO-TEMPORAL ANALYSIS OF MOTOR UNIT ACTION POTENTIAL OF EMG SIGNAL FOR NEUROMUSCULAR DISEASE CLASSIFICATION” submitted by Abul Barkat Mollah Sayeed Ud Doulah, Student No.: 0411062214P, Session: April, 2011 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of MASTER OF SCIENCE IN ELECTRICAL AND ELECTRONIC ENGINEERING on August 31, 2013.

BOARD OF EXAMINERS

1. (Dr. Shaikh Anowarul Fattah)  
Chairman  
Associate Professor  
Department of Electrical and Electronic Engineering  
Bangladesh University of Engineering and Technology  
Dhaka - 1000, Bangladesh.

2. (Dr. Pran Kanai Saha)  
Member  
Professor and Head  
Department of Electrical and Electronic Engineering  
Bangladesh University of Engineering and Technology  
Dhaka - 1000, Bangladesh.

3. (Dr. Mohammed Imamul Hassan Bhuiyan)  
Member  
Associate Professor  
Department of Electrical and Electronic Engineering  
Bangladesh University of Engineering and Technology  
Dhaka - 1000, Bangladesh.

4. (Dr. Farruk Ahmed)  
Member  
Professor  
School of Engineering and Computer Science  
Independent University Bangladesh (IUB)  
Bashundhara, Dhaka-1229, Bangladesh.
CANDIDATE’S DECLARATION

I, do, hereby declare that neither this thesis nor any part of it has been submitted elsewhere for the award of any degree or diploma.

Signature of the Candidate

______________________________
Abul Barkat Mollah Sayeed Ud Doulah
Dedication

To my parents.
Acknowledgments

First and Foremost praise is to ALLAH, the Almighty, the greatest of all, on whom eventually we depend for sustenance and guidance. I would like to thank Almighty Allah for giving me opportunity, determination and strength to do my research. His continuous grace and mercy was with me throughout my life and ever more during the tenure of my research.

I would like to thank my supervisor, Dr. Shaikh Anowarul Fattah, for the patient guidance, encouragement and advice he has provided throughout my time as his student. I have been extremely lucky to have a supervisor who cared so much about my work, and who responded to my questions and queries so promptly. In addition to being an admirable supervisor, he is a man of principles and has immense knowledge of research in general and his subject in particular. I appreciate all his contributions of time, support and ideas. Among many things I learned from Dr. Fattah, is hardworking. I wish to follow one of his good traits just to uphold the spirit “Don’t worry; still we have the whole night to work”.

Part of the thesis work was done in a course project of biomedical signal processing. I would like to thank my group members of that project for their initial contribution.

I would also like to thank the rest of the members of my thesis committee: Prof. Dr. Pran Kanai Saha, Dr. Mohammed Imamul Hassan Bhuiyan, and Prof. Dr. Farruk Ahmed, for their encouragement and insightful comments. I would like to thank the head of the department of Electrical and Electronic Engineering for allowing me to use the lab facilities, which contributed greatly in completing the work in time. I wish to give a special thanks to Dr. Celia Shahnaz, for providing continuous inspiration and guidance to walk the right path of research to success.

Most importantly, none of this would have been possible without the love and patience of my family. I would like to express my heart-felt gratitude to my family.
Abstract

Electromyography (EMG) signal analysis plays a major role in the diagnosis of neuromuscular diseases, such as amyotrophic lateral sclerosis (ALS) and myopathy. With the recorded EMG data generally, either direct analysis is performed or first the motor unit action potentials (MUAPs) are extracted and then MUAP based analysis is carried out. In this thesis, two schemes based on these two approaches are proposed for neuromuscular disease classification. In the direct frame by frame analysis of EMG data, unlike the conventional methods, instead of considering only the local information obtained from a single frame of EMG recording, we propose to utilize global statistics which is obtained based on information collected from some consecutive frames. Different time and frequency domain features are investigated. A discrete wavelet transform (DWT) based feature extraction scheme is developed, where a few high energy DWT coefficients alongwith the maximum value are used, which drastically reduces the feature dimension. With the objective of reducing computational complexity, a MUAP based classification scheme is then presented.

In the proposed MUAP based classification scheme, first MUAPs are extracted from EMG data via EMG decomposition using a template matching scheme. It is well known that not all MUAPs obtained via decomposition are capable of uniquely representing a class. Thus, an energy content based dominant MUAP selection criterion is proposed and only the dominant MUAP is used for the classification. Conventional morphological features of dominant MUAPs are investigated. A feature extraction scheme based on some statistical properties of the DWT coefficients of dominant MUAPs is proposed. Moreover some spectral domain features based on discrete cosine transform are also introduced. For the purpose of classification, the K-nearest neighborhood (KNN) classifier is employed in supervised classification. In order to investigate the performance of the proposed methods, a publicly available clinical EMG database is used. The leave-one-out cross validation technique is used in order to verify the performance in classifying a test data among three classes, normal, ALS and myopathy. It is found that the proposed schemes provide extremely satisfactory results in comparison to that obtained by some of the existing methods in terms of specificity, sensitivity, and overall classification accuracy.
## Contents

Dedication iii

Acknowledgements iv

Abstract v

1 Introduction 1

1.1 Background of EMG signal ................. 1

1.2 History of EMG Signal .................... 2

1.3 EMG Signal Acquisition .................. 3

1.4 Decomposition of EMG Signals ............ 4

1.5 Motor Neuron Diseases .................... 7

1.5.1 Amyotrophic Lateral Sclerosis (ALS) .... 7

1.5.2 Myopathy ............................ 9

1.6 Literature review ........................ 9

1.7 Objective of the Thesis .................. 13

1.8 Organization of the Thesis ............... 13

2 Proposed Raw EMG Signal Based Disease Classification Scheme 15

2.1 Feature Extraction ........................ 15

2.1.1 Preprocessing .......................... 16

2.1.2 Time domain Feature .................. 17

2.1.3 Frequency domain Feature .............. 19

2.1.4 Wavelet Domain Feature ............... 22

2.2 Classifier : k-Nearest Neighbor (KNN) .... 25

2.3 Database Description and Experimental Procedures .... 26

2.3.1 Signal acquisition ..................... 27
3 Proposed MUAP Signal Based Disease Classification Scheme
  3.1 MUAP Recording and Characteristics
    3.1.1 Motor Unit (MU)
    3.1.2 Motor Unit Action Potential (MUAP)
    3.1.3 Morphological Parameters of MUAP
  3.2 EMG Decomposition for MUAP Extraction
  3.3 Proposed Dominant MUAP Selection Method
    3.3.1 Relation between MUAP energy and motor unit characteristics
    3.3.2 Energy content based dominant MUAP selection
  3.4 Proposed DCT Based Feature Extraction
  3.5 Proposed DWT Based Feature
  3.6 Results and Analysis
  3.7 Conclusion

4 Conclusion
  4.1 Concluding Remarks
  4.2 Major Contribution of This Thesis
  4.3 Scope and Future Work

A EMGLAB: EMG Decomposition Method
List of Tables

2.1 Performance of proposed classification method experimented on two
class small dataset ............................................. 33
2.2 Performance of proposed classification method experimented on two
class large dataset ............................................ 34
2.3 Performance of proposed classification method experimented on three
class small dataset ............................................. 35
2.4 Performance of proposed classification method experimented on three
class large dataset ............................................ 36

3.1 Performance of proposed classification method experimented on two
class small dataset ............................................. 66
3.2 Performance of proposed classification method experimented on two
class large dataset ............................................ 67
3.3 Performance of proposed classification method experimented on three
class small dataset ............................................. 68
3.4 Performance of proposed classification method experimented on three
class large dataset ............................................ 69
List of Figures

1.1 EMG data pattern of a normal person, the ALS patient and the myopathic patient ......................................................... 2
1.2 Surface EMG electrodes and its placement .......................... 4
1.3 Different indwelling needle electrode types. A) In the monopolar design, the needle is Teflon coated and its exposed tip serves as the active electrode. B) The concentric needle has an active electrode running as a thin wire through the needle center and exposed at the tip. The cannula serves as reference. C) The quadrifilar needle electrode is 25-gauge stainless steel. The needle electrode was custom designed with four 50µm platinum wire electrodes equably exposed on the side port of the cannula ......................................................... 5
1.4 EMG signal decomposition is the process of discovering the significant constituent MUAPTs that contribute to a detected EMG signal .... 6
1.5 The normal and affected nerve cell of human body .................. 8
1.6 Muscular dystrophy in biceps muscle .................................. 9
2.1 RMS values of EMG data for normal person, myopathy and ALS patients ............................................................... 17
2.2 Prediction error filter \( A(z) = 1 + a_1 z^{-1} + \ldots + a_p z^{-p} \). The output of this filter is approximately white noise if the prediction order is large enough. .......................................................... 21
2.3 Sub-band decomposition of DWT implementation; \( h[n] \) is the high pass filter, \( g[n] \) the low pass filter ................................. 24
2.4 Representation of data points being classified by k-Nearest Neighbors classifier. The parameter \( k \) equals a value of 3 and the point takes the label of the class that is in majority which is the red one ....... 26
2.5 Proposed maximum DWT feature for normal persons and ALS patients (Two class small dataset) ........................................... 31
2.6 Proposed maximum DWT feature for normal persons and ALS patients (Two class large dataset) .................................................. 32
2.7 Proposed maximum DWT feature for normal, myopathic and ALS patients (Three class small dataset) .................................................. 32
2.8 Proposed maximum DWT feature for normal, myopathic and ALS patients (Three class large dataset) .................................................. 33

3.1 Illustration of the MUAP parameters ............................................. 40
3.2 Main types of pathology in the motor unit. Schematic 3-D and cross-sectional illustrations of motor units as well as the corresponding EMG characteristics. (A) Two normal overlapping MUs, (B) Neuropathic motor unit with enlargement of one motor unit and some denervated muscle fibres. Large MUAP (C) Myopathic motor units. Small MUAP ................................................................. 43
3.3 MUAPs from a normal subject, a patient with neuropathy and a patient with myopathy ................................................................. 44
3.4 EMGLAB computer screen analyzing an EMG signal .................. 46
3.5 Template panel, showing the templates of the identified MUAPs .... 47
3.6 MUAPs after decomposition of single normal EMG signal .......... 49
3.7 Energy content of extracted MUAPs from single normal EMG signal 50
3.8 MUAPs after decomposition of single ALS EMG signal .............. 51
3.9 Energy content of extracted MUAPs from single ALS EMG signal .. 52
3.10 MUAPs after decomposition of single myopathy EMG signal ....... 52
3.11 Energy content of extracted MUAPs from single myopathy EMG signal 53
3.12 Absolute DCT coefficients of dominant MUAP extracted from a normal person ................................................................. 53
3.13 Absolute DCT coefficients of dominant MUAP extracted from a ALS patient ................................................................. 54
3.14 Absolute DCT coefficients of dominant MUAP extracted from a myopathy patient ................................................................. 54
3.15 Mean values of absolute DCT coefficients of dominant MUAPs obtained from different EMG recordings corresponding to normal, ALS and myopathy subjects .............................. 55
3.16 Maximum values of absolute DCT coefficients of dominant MUAPs obtained from different EMG recordings corresponding to normal, ALS and myopathy subjects .............................. 55
3.17 Wavelet decomposition of a normal dominant MUAP ....................... 57
3.18 Wavelet decomposition of a ALS affected dominant MUAP .............. 58
3.19 Wavelet decomposition of a myopathy affected dominant MUAP ....... 59
3.20 Multiple level wavelet decomposition of dominant MUAP ............... 60
3.21 Proposed DWT feature for normal persons and ALS patients (Two class small dataset) ................................................................. 61
3.22 Proposed DWT feature for normal persons and ALS patients (Two class large dataset) ................................................................. 62
3.23 Proposed DWT feature for normal, myopathic and ALS patients (Three class small dataset) ......................................................... 63
3.24 Proposed DWT feature for normal, myopathic and ALS patients (Three class large dataset) ......................................................... 64
3.25 Effect of variation of decomposition levels on total accuracy for both two class small dataset (top) and large dataset (bottom) ............ 65
3.26 Effect of variation of decomposition levels on total accuracy for both three class small dataset (top) and large dataset (bottom) ............ 65
3.27 Classification performance of two class small dataset ..................... 70
3.28 Classification performance of two class large dataset .................... 70
3.29 Classification performance of three class small dataset .................. 71
3.30 Classification performance of three class large dataset .................. 71

A.1 Signal panel, showing a segment of the EMG signal (top trace) and the residual signal after template matching (bottom trace) ............ 76
A.2 Template panel, showing the templates of the identified MUAPs ........ 77
A.3 Discharge panel, showing the identified discharge times of each MU. The vertical lines indicate the interval displayed in the signal panel . 78
A.4 Close-up panel, showing a segment of the EMG signal at an expanded scale (top trace), a configuration of templates selected to match it (bottom traces and numbers), the sum of the templates (top dotted trace), and the residual (center trace). All the time scales are in seconds from the start of the signal.
Chapter 1

Introduction

1.1 Background of EMG signal

The electromyography (EMG) signal is a biomedical signal that measures electrical currents generated in muscles during its contraction representing neuromuscular activities. It is complicated in nature as the muscle activity (contraction/relaxation) is continuously controlled by the nervous system. The EMG signal is dependent on the anatomical and physiological properties of muscles. It serves as a reliable source of information about different features of muscle function.

There are two types of EMG, namely clinical (diagnostic) and kinesiological. Clinical EMG focuses on characterizing motor unit action potentials (MUAPs) during muscle contraction. A motor unit is a single alpha motor neuron and the muscle fibers it activates. Kinesiological EMG investigates the relationship of muscular function to movement of the body segments and evaluates timing of muscle activity with regard to the movements. Clinical EMG is carried out to help diagnose neuromuscular disorders, while kinesiological EMG mostly deals with human movement analysis.

Recording of EMG signals can be extracted by placing electrodes on the surface of the skin (surface EMG) or inserting fine wire or needle electrodes deep (approximately 0.25 – 0.5cm) inside the muscle (indwelling/needle EMG). An advantage of needle EMG is its selectivity which strongly attenuates cross talk (a phenomenon in which MUAPs from muscles further away also contribute to the recording through volume conduction) and can be used when targeting small muscles. In this thesis, EMG shall refer to clinical EMG signals recorded within the muscle (indwelling).

Typical EMG data patterns of a normal person, an ALS patient and a myopathic
Fig. 1.1: EMG data pattern of a normal person, the ALS patient and the myopathic patient are shown in Fig. 1.1.

### 1.2 History of EMG Signal

The study of EMG originated from the discovery of the role that electricity plays in the nervous system of animals. In the mid 1600’s, it was shown that a highly specialized muscle was the source of an electric ray fish’s energy. Direct observation of the relationship between muscles and electricity was performed in dissected frogs, by Luigi Galvani (1791-1797) [2]. Following Galvani’s work, Carlo Matteucci demonstrated that cell membranes had a voltage across them and could produce direct current. German physiologist Emil du Bois-Reymond, who was inspired by Matteucci’s work, went on to discover the action potential in 1848. The conduction velocity of action potentials was first measured in 1850 by du Bois-Reymond’s friend, Hermann von Helmholtz. In 1880, Wollaston hypothesized that a sound is generated during muscle contraction and demonstrated that energy generated by muscles was in the audible range [3]. EMG signals were first displayed on an oscilloscope in the 1920’s. This knowledge has increased the curiosity among several investigators in the field of electromyography.
1.3 EMG Signal Acquisition

The EMG signal gets contaminated with undesirable signals or noise (e.g. movement artifact, 60 Hz artifact, etc.) as it traverses from its source to the recording apparatus, making it difficult to detect. With regards to recording the EMG signal, the amplitude of the MUAP depends on many factors which include: diameter of the muscle fiber, distance between active muscle fiber and the detection site (adipose tissue thickness), and filtering properties of the electrodes themselves. There are differential amplification based techniques which can be used to record EMG signals reliably with high signal to noise ratio (SNR). Care should be taken when selecting electrodes and type of amplifiers. Filtering is normally performed so that unwanted signals, such as low frequency motion artifacts are removed from the EMG signal.

As mentioned before, there are two main types of electrodes for EMG data acquisition: surface and fine wire/indwelling. The surface electrodes are of two types: active and passive. The active surface electrode has built-in amplifiers at the electrode site. For this type of electrodes no gel is required and they provide high SNR with reduced movement artifacts. The passive surface electrode detects the EMG signal without a built-in amplifier, making it important to reduce all possible skin resistance as much as possible. It requires conducting gels and extensive skin preparation. With passive electrodes, the SNR decreases and many movement artifacts are amplified along with the actual signal once amplification occurs. The advantages of surface electrodes are that there is minimal pain with application, they are more reproducible, easy to apply, and good for movement applications. The disadvantages of surface electrodes are that they have a large pick-up area and therefore, have more potential for cross talk from adjacent muscles. Additionally, these electrodes can only be used for surface muscles. Fig. 1.2 illustrates some of these EMG electrodes and its placement in hand.

Fine wire electrodes require a needle for insertion into the muscle. The advantages of fine wire electrodes are an increased SNR, a more specific pick-up area, ability to test deep muscles, isolation of specific muscle parts of large muscles, and ability to test small muscles which would be impossible to detect with a surface electrode due to cross-talk. The disadvantages are that the needle insertion causes discomfort, the uncomfortableness can increase the tightness in the muscles, cramp-
Fig. 1.2: Surface EMG electrodes and its placement

ing may occur, the electrodes are less repeatable as it is very difficult to place the needle/fine wires in the same area of the muscle each time. Additionally, one should stimulate the fine wires to be able to determine their location, which increased the un-comfortableness of using this type of electrode. However, for certain muscles, fine wires are the only possibility for obtaining their information. The most common indwelling electrodes used to record EMG are monopolar and concentric needle electrodes. In addition there are other EMG recording techniques based on special needle electrode configurations, namely: single-fiber EMG, macro EMG and quadrifilar needle electrode. Fig. 1.3 illustrates some of these EMG needle types.

The voltage level corresponding to the electrical activity of the muscle is usually very low in magnitude (generally less than $5 \times 10^{-3}$ mV). It is therefore essential to amplify the detected EMG signal so that the signal can be sampled reliably by typical analog to digital converters. The most practical way of amplifying EMG signals is by using an instrumentation amplifier. It is essential that the EMG signal is amplified without distortion.

### 1.4 Decomposition of EMG Signals

Quantitative analysis of EMG signals recorded using needle electrodes can provide important information for clinical diagnosis of neuromuscular disorders. For the purpose of analysis, either one can directly use the extracted EMG signal or can decompose the recorded signal into its constituent MUAP trains (MUAPT). An
MUAPT is a set of action potentials that represent a given alpha neuron and the muscle fibers it innervates (motor unit). It can provide information regarding the way motor units behave during muscle contraction, which helps in the diagnosis of various neuromuscular disorders. Fig. 1.4, conceptually presents the process of EMG signal decomposition and depicts the relationship between a decomposed EMG signal and the activity of individual motor units [4].

Classical way of assessing of analyzing MUAPs via waveshape observation or listening to audio characteristics may not suffice when it comes to detecting disorders that are not readily apparent. Now-a-days EMG signals are decomposed into their individual MUAPTs using software routines with manual supervision. The EMG decomposition process generally consists of four stages; pre-processing, detection, classification and resolution of superimposed waveforms. Pre-processing includes high-pass filtering (1000 Hz) and low-pass differentiators to reduce the probability
EMG signal decomposition is the process of discovering the significant constituent MUAPTs that contribute to a detected EMG signal of temporal overlap among MUAPs and therefore reduce the number of superimposed waveforms. Detection of MUAPs is generally performed by using slope and amplitude thresholds. During the classification stage, MUAPs are grouped into different classes (motor units). The most widely used method in identifying individual MUAPs from a composite EMG signal is known as template matching. Supervised classification via a train set is also employed [5], [6], [7]. Difficulty in classification comes when two or more action potentials superimpose on one another. Superimposing occurs because some motor units fire at (almost) the same time, which means that the detected potential is an algebraic sum of individual motor unit potentials from different motor units. In superimposed waveforms, peaks of overlapping MUAPs may be distinguishable or the peaks of individual MUAPs may combine to make one large peak or even out-of-phase peaks are summed together and end up canceling out one another.

There are a number of algorithms for resolving superimposed MUAPs [6]. These
algorithms are based on two approaches. First one is known as the peel-off or sequential approach. It is based on matching MUAPs, one at a time, with the superimposed waveform. Once a match is suitable, the MUAP is subtracted from it and the created residual waveform is then used to search for other contributing MUAPs. In [8], an efficient ‘peel-off’ technique base EMG decomposition toolbox, namely EMGLAB is introduced. In [9], spectrum matching in the wavelet domain is used for EMG decomposition. The second method of resolving superpositions is based on synthesizing superimposed waveform models by adding up combinations of MUAPT templates with different relative time shifts. Model synthesis and comparison is repeated until an optimal or acceptable match is obtained between one of the model superimposed waveforms and the actual superimposed MUAP of time shifts considered for matching. This type of EMG decomposition methods generally limit the number of assumed contributing MUAPs to 2 or 3 [6]. Some of the EMG decomposition techniques to be mentioned are knowledge-based processing [10], [7], neuron-muscle communication based scheme [11], NNERVE (neural network extraction of repetitive vectors for electromyography) [5], wavelet based multichannel scheme [12], and fuzzy logic based system [13].

1.5 Motor Neuron Diseases

Motor neuron diseases (MNDs) are a group of disorders of the nervous system responsible for the progressive degeneration of motor neurons [14]. Among different motor neuron diseases, two most common diseases are amyotrophic lateral sclerosis (ALS) and myopathy, which are briefly discussed below.

1.5.1 Amyotrophic Lateral Sclerosis (ALS)

ALS is the most common variant of motor neuron diseases [14]. It is a progressive neurodegenerative disorder that affects both the upper and lower motor neurons. It can occur among young individuals, but it most commonly affects people between the ages of 40 – 70, with a slight male predominance. It can be difficult to diagnose in the early stages because its symptoms may mimic other disorders. However, there are clinical signs that can be indicative of a wasting of either the upper or lower motor neurons. A lower motor neuron lesion is characterized by muscle atrophy, weakness,
fasciculations and cramps. An upper motor neuron lesion can manifest as stiffness, spasticity, weakness, pathologic hyperreflexia, and Babinski responses. Five to ten percent of people with ALS have a hereditary form of the disease whereas 90 to 95 percent of ALS is sporadic. Over time, as the disorder progresses, ALS robs those affected of most physical activity. It destroys motor neurons that are vital in the functioning of the nervous system, and through which the brain controls voluntary muscles in the body. As a result, leg and foot muscles (controlled by motor neurons in the lower spinal cord), arm, hand and finger muscles (controlled by motor neurons in the upper spinal cord), speaking, swallowing and chewing (controlled by motor neurons in the brain stem) are progressively affected. The normal and affected nerve cell of human body is depicted in Fig. 1.5. Life expectancy of newly diagnosed people is roughly three to five years. Over time, ALS may spread throughout the body and may affect muscles required for breathing. To date, despite years of research, the cause of ALS has not been uniquely determined, making the search for a cure very difficult [15].

![Fig. 1.5: The normal and affected nerve cell of human body](image-url)
1.5.2 Myopathy

In general, myopathy is the disease of muscles. Muscle injury, infection, inherited disorders affecting muscle functions and thyroid diseases may cause myopathy. Myopathies which are caused by inherited genetic defects, by endocrine inflammatory and by immune inflammation are known as muscular dystrophies, polymyositis and dermatomyositis, respectively. Weakness of the large muscles (muscles of the center- proximal muscles) around the neck, shoulders and hip are the symptoms of myopathies observed frequently. Myopathy can be inherited from birth, but it is found mostly between the ages of 20 to 50. Women have chances of being infected by myopathy twice than the men. Sometimes regular therapy minimizes the severity of the disease. Again if less responsive to the treatment, especially in the case of inherited myopathies, condition of the patient worsens with time [15], [16], [17]. A diseased muscle’s pattern is shown in Fig. 1.6.

![Muscular dystrophy in Biceps muscle](image)

Fig. 1.6: Muscular dystrophy in biceps muscle

1.6 Literature review

Neuromuscular disorders are a sort of disease that affects the nervous system consists of the brain, spinal cord, sensory organs and muscles. Early detection and proper diagnosis of these diseases is of prior requirement for living. Clinical examination of EMG signals is one of the best options to reveal important characteristics about
these diseases and which may disclose the nature and eventual treatment of these diseases [18]. There are two basic ways of analyzing the EMG signal: 1) raw EMG signal analysis 2) EMG decomposition and its functional unit; i.e. Motor Unit Action Potential (MUAP) classification.

Many researches have been done in the past to extract valuable information from EMG signal. Researches are mainly concentrated on quantitative analysis of raw EMG signal and quantitative needle EMG analysis consisting of two major processes: EMG decomposition and MUAP classification.

EMG decomposition is the process of breaking down to its constituent’s action potentials. In [19], EMG decomposition method based on photographically recorded MUAPs is developed. A special three channel recording electrode and a visual-computer decomposition scheme based on template matching and firing statistics for MUAP identification is proposed in [20]. In [21], their original system used waveform template matching, whereas more recently [22] they used different shape parameters as input to a template matching technique. A combination of shape recognition of the MUAPs and statistical probability of occurrence is utilized in [23]. In [9], they developed a comprehensive technique to identify single MU potentials based on one-channel EMG recordings measuring waveform similarity of single MU potentials in the wavelet domain. Recently, an EMG signal decomposition technique using motor unit potential train validity is proposed in [24].

Classification of raw EMG signal has also been done with notable accuracy. Most of the methods analyze the raw EMG signal on frame by frame basis. The frames are then considered as class representative. In order to gain quantitative information from EMG signal, different time domain, frequency domain and time-frequency domain approaches have been extensively investigated. Time domain feature sets are generally extracted directly using the EMG signal or its frame and then employed in different pattern classifier. In [25], they compared the number of zero crossings with force production and found a linear relationship between various parameters of the signal and the forces generated in the muscle up to 50% of a maximum voluntary contraction. The median frequency increases with force up to a point equal to 50% of the maximum contraction is shown in [26]. Visual inspection of the raw EMG signal together with appropriate graphics of the joint angles is proposed
In [27], the author found that the turns/amplitude analysis method was more useful than other methods. It is reported in [28], that median frequency deviates from its normal value in a number of neuromuscular disorders. Among the newest methods of EMG analysis are those involving wavelet analysis, which examines both the frequency and time domain combined. It is demonstrated in [29] that how the CWT is mathematically similar to surface EMG signals with noise and is therefore the favorite candidate for analyzing these signals. The wavelet transform method has been applied with increased accuracy. Recently, a autoregressive (AR) and wavelet transform based approach for EMG signal classification is proposed in [30].

In MUAP based EMG signal analyses, the areas of interests are either in the MUAP decomposition accuracy or MUAP based disease classification. The MUAPs are extracted either in a direct manner from single fiber needle electrode insertion [31] [32] [33] or decomposition of EMG recording [34] [35]. Different domain features are then extracted from all MUAPs. Most of MUAPs are generally taken from biceps brachii muscles and the number of total patients vary from 10—50 in the literature. Typically, the literature discusses how feature values for a set of MUAPs detected from to characterize as myopathic, normal or neuropathic. These three categories provide an initial step towards a useful, robust neuromuscular clinical decision support systems. In order to achieve feature from MUAP, both morphological features [34] and statistical features [18] are obtained for analysis.

In [34] used 7 morphological features (duration, spike duration, amplitude, area, spike area, number of phases and number of turns) of the MUAPs to classify normal patients,neuropathic and myopathic patients . In this work, the parametric pattern recognition (PPR) algorithm that facilitates automatic MUAP feature extraction and Artificial Neural Network (ANN) models are combined for providing an integrated system for the diagnosis of neuromuscular disorders.

In [31] the proposed method uses Fisher’s Linear Discriminant Analysis (LDA) to determine the probability of the abnormality of each MUP (i.e. MUAP characterization). The technique estimated the conditional probabilities (one for each category) of a MUAP being detected from a muscle with a given category of disorder. Duration, area, turns count and center frequency are used as the features. The error rates using a single MUP for muscle characterization are 52.9% for my-
pathic disorders, 44.2% for neuropathic disorders and 35.2% for normal muscles. The major limitation of this work is the high error rate.

In [32], a computer based system for acquiring MUAPs is developed and evaluated. The mean and standard deviations of amplitude, area and thickness (area-to-amplitude ratio) values for sets of MUAPs are used as features. The method classifies a muscle as myopathic if one or more of the mean feature values falls below the normal range and classifies the muscle as neuropathic if one or more of the features falls above the normal range. The method is tested with MUAPs detected from 50 patients with known myopathic and 55 patients with known neuropathic disorders.

In [33], they developed an outlier method for characterization of diseases. The conventional morphological MUAP parameters are taken as features. Since the outlier method is as sensitive as mean values, some misleading MUAPs may affect the performance.

In [35], the authors proposed a method that automatically decomposes the recorded EMG signal to MUAPs and classifies using morphological MUAP feature. The preprocessing of the EMG recordings resulted in 9919 candidate MUAPs and the proposed method evaluates 365 template MUAPs under the suggestion of experienced neurophysiologist; which imposes a complete dependence on human interaction. Finally, the classification accuracy of 89% is reported using Radial Basis Function Artificial Neural Networks and decision trees.

In [18], the same features of MUAPs are taken as their previous work in [34]. In addition to that, wavelet based statistical features are proposed with less classification accuracy. Apart from the accuracy, the proposed method restricts the MUAP feature extraction from some specific detail coefficients of 6-level DWT decomposition, which may not be feasible and realizable for all MUAPs.

A common drawback of the conventional direct EMG signal analysis is to consider all frames as individual class representative. The single frame may not carry sufficient information. Hence a method is still demanding in which the global information are extracted from continuous selected frames instead of single frame.

MUAP based methods consider extracting features from all available MUAPs. However, there is always some non-stationary MUAPs in EMG signal which may pro-
vide misleading information and decreases the quality of extracted features. Therefore, a robust scheme is required for selecting the MUAPs that carry significant information for neuromuscular disease classification.

1.7 Objective of the Thesis

The objectives of this thesis are to:

1. Extract time and frequency domain features from the raw EMG signal.
2. Analyze MUAPs obtained from EMG decomposition.
3. Develop a feature extraction scheme based on dominant MUAP.
4. Investigate the effect of feature dimension reduction on classification accuracy.
5. Develop efficient classification scheme for the diagnosis of ALS, myopathy and normal control group.

1.8 Organization of the Thesis

The proposed method involves two different approaches for neuromuscular disease classification: (1) based on raw EMG signal and (2) based on dominant MUAP obtained by EMG signal decomposition.

In Chapter 2, raw signal based neuromuscular disease classification method is described. Overall there are four major steps involved in this method: preprocessing, frame selection, feature extraction from frames of EMG signal and classification. Prior to frame selection, preprocessing by filtering is done on the raw EMG signal. After that frame selection scheme from EMG signal is presented. In the later section, different time domain, frequency domain and time-frequency domain features are extracted from the frames of EMG signal. A wavelet based feature is then proposed and described. Finally experimental results on publicly available clinical EMG database are presented along with comparative performance analysis.

In Chapter 3, the proposed dominant MUAP based classification scheme is presented which involves four major tasks, such as EMG signal decomposition, dominant MUAP selection, feature extraction and classification. First, the EMG signal
is decomposed by an open source widely used decomposition tool ‘EMGLAB’. Next, the proposed dominant MUAP selection method is presented. In the next section, the dominant MUAPs are considered for feature extraction. Different morphological, frequency domain and proposed wavelet domain features are then extracted. Finally experimental results with comparative analysis are demonstrated considering the same database.

Chapter 4 summarizes the outcome this thesis with some concluding remarks and possible future works.
Chapter 2

Proposed Raw EMG Signal Based Disease Classification Scheme

A precise and computationally efficient means of classifying EMG signal patterns has been the subject of considerable research effort in recent years. Quantitative analysis of recorded EMG signal provides an important source of information for the diagnosis of neuromuscular disorders. With the assistance of the recent development of computer-aided EMG equipment, different methodologies in the time domain, frequency domain and time-frequency domain have been followed for quantitative analysis. The raw EMG signal based method mainly concentrated on four steps: preprocessing, frame selection, feature extraction from frames of EMG signal and classification. Prior to frame selection, preprocessing by filtering is done on the raw EMG signal. Following the step, frame selection scheme from EMG signal is presented. In the later section, different time domain, frequency domain and time-frequency domain features are extracted from the frames of EMG signal. A wavelet based feature is then proposed and reported. Finally experimental results on publicly available clinical EMG database are presented along with comparative performance analysis.

2.1 Feature Extraction

The purpose of including feature extraction algorithms in a pattern recognition systems is to gather a set of information that efficiently and optimally represents significant characteristics from the input data. One of the objectives of this research is to develop unique feature extraction algorithm which can reduce the overall data amount while retaining or enhancing signal characteristics inherent to each class
in order to improve classification performance. In view of this some of the existing
features are investigated and some new features are proposed with a target to develop
an accurate neuromuscular disease classification scheme.

Feature can be acquired directly from time-segments of the sampled recorded
data without any preprocessing. However, using appropriate preprocessing steps for
extracting features improves the recognition rate and the generalization capability
of the classifier. Therefore, the preprocessing of recorded EMG signal is first per-
formed. With the preprocessed EMG data, feature extraction is carried out in time,
frequency, and time-frequency domains.

2.1.1 Preprocessing

Time variation pattern of the EMG signal is very complicated in nature and thus
it would not be a convincing approach to classify them directly based on the time
variation of the data as observed. As a result, further detailed analysis using both
temporal and spectral representations would be definitely helpful in EMG data clas-
sification. Instead of working with the whole duration of the recorded EMG data,
generally some portions of the data are selected. For selecting the significant portion
of the EMG data, the energy content in different frames is considered. For a frame
of EMG data \( x(n) \) containing \( N \) samples the root mean square (RMS) value can be
obtained as

\[
E_i = \sqrt{\frac{1}{N} \sum_{n=0}^{N-1} x^2(n)}
\]

(2.1)

where \( i \) indicates the frame number. In order to demonstrate the energy variation
in different frames, in Fig. 2.1, the RMS values obtained from each non-overlapping
frame of the EMG data considering two normal persons, two myopathy patients
and two ALS patients are shown. It is found that the RMS values corresponding
to the ALS patients fluctuate abruptly in the initial and final frames but exhibit a
stable range of values in the middle portion of the whole recording. On the other
hand, RMS values corresponding to a normal person and myopathy patient exhibit
a relatively steady range of RMS values all over the duration of the recorded data.
For a fair comparison, in the proposed method, only the data extracted from the
middle portion of the recording are taken in consideration for both normal and
diseased cases. Since the energy of the EMG signal is mostly concentrated in the
low frequency regions, a low pass filter is used to reduce the effect of high frequency regions. The low pass filtered EMG signal is then used for feature extraction.

2.1.2 Time domain Feature

Different time domain features are most commonly used for the analysis of EMG data. In most of the research works, different morphological features are used, such as amplitude, pitch, integrated EMG, mean absolute value, EMG variance, zero crossing, Willison amplitude etc \[36\]. One major advantage of time domain features is the ease of computation. Moreover these features can easily be interpreted with respect to the observed temporal variation. The conventional time domain features are investigated in this research and results obtained by using those features are also reported. However, in what follows only the time domain features those are used in the proposed scheme are described.

**Autocorrelation**

The cross-correlation between two signals is a measure of dependency of these two signals on each other. Higher the dependency, larger will be the value of cross-correlation. When the two signals involved in the cross-correlation operation become exactly same, the operation is then termed as autocorrelation. In fact, an autocorrelation sequence reflects the degree of similarity at different portions of a time series.
data. Hence it is a well known operation for measuring the hidden periodicity of a signal [37].

For an $N$-length sequence of EMG data $x(n)$, its autocorrelation function $r_x(\tau)$ can be computed as

$$r_x(\tau) = \frac{1}{N} \sum_{n=0}^{N-1} x(n)x(n+|\tau|),$$

(2.2)

where $\tau$ denotes the correlation lag. In this research, the characteristics of the autocorrelation function of different frames of EMG data have been investigated. It is found that no consistent information regarding the hidden periodicity of the EMG signal is readily observable. Hence, correlation based periodicity or pitch measures have not taken into consideration as potential features. Magnitude of the zero lag of the autocorrelation function of a frame, which in fact represents the frame energy, exhibits relatively better feature quality in terms of within class compactness and between class separation. If a higher dimension of feature vector containing first few lags is considered, the quality of feature degrades. Hence, zero lag of the autocorrelation function of a frame is chosen as feature.

**Zero-crossing rate (ZCR)**

The zero-crossing rate (ZCR) expresses the number of times a signal crosses the axis of abscissas. It can be defined as

$$Z_i = \frac{1}{2N} \sum_{n=1}^{N-1} |\text{sgn}[x(n)] - \text{sgn}[x(n-1)]|$$

(2.3)

where

$$\text{sgn}(x) = \begin{cases} 
1, & x \geq 0 \\
-1, & x < 0 
\end{cases}$$

(2.4)

The random temporal fluctuations of the EMG signal may serve as distinguishable feature. It is observed that for a normal person, the value of ZCR in different frames is considerably consistent and the ZCR value increases for diseased person. However, the same observation could not attained for other large datasets. Hence, the ZCR is not a consistent distinguishable feature to comment on the detection of diseases.
2.1.3 Frequency domain Feature

One interesting property of the EMG signal is its audible frequency content which inspires many researchers to investigate the frequency domain characteristics of the EMG data. The main advantage of frequency domain analysis over the time representation is that it allows a clear visualization of the periodicity of the signal along with energy concentration in different frequency ranges. In what follows, the frequency domain features those are used in the proposed classification scheme are described.

Spectral Magnitude

Let us consider a frame of EMG data consists of $N$ discrete values $x(n)_{n=0}^{N-1}$ sampled at a time interval of $\Delta$. The discrete Fourier transform (DFT) of this signal is defined as

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-i2\pi kn/N}, \quad k = 0, \ldots, N - 1$$ (2.5)

and corresponding inverse DFT is given by

$$x(n) = \frac{1}{N} \sum_{k=0}^{N-1} X(k) e^{i2\pi kn/N}, \quad n = 0, \ldots, N - 1$$ (2.6)

Note that the DFT gives $N/2$ independent complex coefficients, thus giving a total of $N$ values as in the original signal and therefore being nonredundant. Here the frequency resolution will be $\Delta f = \frac{1}{N\Delta}$.

In order to investigate the spectral characteristics of the EMG signal only magnitude spectrum of the EMG signal is taken into consideration. In particular, computing the FFT of a frame of EMG data, only the magnitude of the extracted FFT coefficients are taken into consideration. It is expected that within a short duration of the EMG data, the spectral behaviour remains consistent. Hence from a long duration of the EMG recording, for short time spectral analysis, smaller frames are extracted by using windowing techniques. However, effect of windowing in time domain may generate unwanted ripples in spectral domain. It is observed from extensive experimentation on different types of EMG recordings that the maximum peaks of the magnitude spectra for different frames appear at different frequency locations in a random fashion. As a result, instead of considering the frequency
values corresponding to the maximum spectral peaks, only their amplitude values are taken into consideration as a spectral feature.

**Parametric Power Spectral Density**

Since the EMG spectrum consists of some peaks, the autoregressive (AR) or all-pole modeling is most suitable to fit the EMG data in comparison to the other two models, namely the moving average (MA) or all-zero model and the autoregressive moving average (ARMA) or pole-zero model. To identify the AR model parameters $a(k)$, the theory of linear prediction is most widely used where a linear combination of the previous samples of a signal is used to predict the current sample. For the prediction of an AR sequence at the $n$-the instant

$$x_p(n) = \sum_{i=1}^{p} -a_i x(n - i)$$  \hspace{1cm} (2.7)

where $p$ is the prediction order indicating that only the $p$ previous values of the sequence are used in the estimation and $a_i$ are the prediction coefficients. Defining $a_0 \equiv 1$, the prediction error can be written as

$$e(n) = x(n) - x_p(n) = \sum_{i=0}^{p} a_i x(n - i)$$  \hspace{1cm} (2.8)

The output of the prediction error filter with a transfer function $A(z) = 1 + a_1 z^{-1} + \ldots + a_p z^{-p}$ is approximately a white noise process if the prediction order is large enough.

The transfer function of the AR system is $H(z) = 1/A(z)$ and the AR system output can be expressed as

$$x[n] = -\sum_{i=1}^{p} a_i x[n - i] + e[n].$$  \hspace{1cm} (2.9)

Several methods exist to estimate the autoregressive parameters, such as least squares, Yule-Walker and Burgs method [38]. These estimation techniques lead to approximately the same parameter estimates [39]. In this study, the most widely used Yule-Walker method is employed to estimate the AR coefficients, where starting from (2.9) following matrix equation is derived and the AR parameters are obtained via matrix inversion.
Fig. 2.2: Prediction error filter \( A(z) = 1 + a_1 z^{-1} + \ldots + a_p z^{-p} \). The output of this filter is approximately white noise if the prediction order is large enough.

\[
\begin{bmatrix}
  r(0) & r(1) & r(2) & \cdots & r(p) \\
  r(1) & r(0) & r(1) & \cdots & r(p-1) \\
  r(2) & r(1) & r(0) & r(p-2) & \vdots \\
  \vdots & \vdots & \vdots & \ddots & \vdots \\
  r(p) & r(p-1) & r(p-2) & \cdots & r(0)
\end{bmatrix}
\begin{bmatrix}
  1 \\
  a_1 \\
  a_2 \\
  \vdots \\
  a_p
\end{bmatrix}
= \begin{bmatrix}
  E \\
  0 \\
  0 \\
  \vdots \\
  0
\end{bmatrix}
\quad (2.10)
\]

Here \( E \) is the sum of the squared errors and \( r(i) \) denotes the \( i \)th lag of the autocorrelation function which can be computed by using (2.2). The coefficients \( a_1 \) to \( a_p \) and the error term \( E \) can be obtained by solving the \( P+1 \) equations with \( P+1 \) unknowns represented by the above matrix equation. Once the AR parameters are computed, the parametric PSD can be obtained as

\[
\hat{P}_x(f) = \frac{E}{|1 + \sum_{k=1}^{p} \hat{a}_p(k)e^{-j2\pi fk}|^2}
\quad (2.11)
\]

The choice of model order is an important issue in case of dealing with biomedical signal, especially the EMG, which is very complex in nature. The optimum model order is best estimated by maximizing the goodness of fit, that is, minimizing \( E \) in (2.10). There are various criteria proposed for choosing the best model order, such as Akaike’s information-theoretic criteria, Parzen’s criterion, and Schwartz and Rissanens minimum description length [39]. Based on experimental analysis on different EMG signals, it is found that instead of using variable model order obtained
by using a conventional criterion, it is sufficient to consider a fixed high order (e.g. 15). Once the AR parameters of such a predefined order are estimated, the AR PSD is then computed and utilized directly in the classifier.

**Median frequency**

It is well known that the muscle fatigue results in a downward shift of frequency spectrum of the EMG signal. Hence, it is expected that the use of some statistical measures of the extracted frequency components of the EMG signal can provide useful insight into the nature of the EMG data. Among such measures, mean frequency (MNF) and median frequency (MDF) are used in the proposed method. The MNF is defined as the mathematical mean of the spectrum curve, i.e.

\[
MNF = \frac{\sum_{j=1}^{M} f_j P_j}{\sum_{j=1}^{M} P_j}
\]  

(2.12)

where \(f_j\) is the frequency value of EMG power spectrum at the frequency bin \(j\), \(P_j\) is the EMG power spectrum at the frequency bin \(j\), and \(M\) is the length of frequency bin. The MDF is defined as a half of the total power, or TTP (dividing the total power area into two equal parts). The definition of MDF is given by

\[
\sum_{j=1}^{MDF} P_i = \sum_{j=MDF}^{M} P_j = \frac{1}{2} \sum_{j=1}^{M} P_j
\]  

(2.13)

The behaviour of MNF and MDF is almost similar. However, the performance of MNF in each of the applications is quite different compared to the performance of MDF. In this research, only the MDF feature extracted directly from recorded EMG signal is used for neuromuscular disease classification.

**2.1.4 Wavelet Domain Feature**

Apart from time domain and frequency domain analyses, there are some signal analysis techniques that offer simultaneous interpretation of the signal in both time and frequency domains. Among several time-frequency domain methods, wavelet transform has emerged as the most favoured tool by researchers as it provides sufficient information both for analysis and synthesis of the original signal along with a significant reduction in the computation time.
Discrete wavelet transform (DWT)

The discrete Wavelet transform (DWT) is one of the most popular time-frequency representation techniques, which is capable of representing a signal into a two-dimensional function of time and frequency. It offers several advantages in comparison to the conventional frequency transforms, such as Fourier transform and cosine transform. For example, in short time Fourier transform analysis, a constant resolution is used at all frequencies but the DWT being a multi-resolution technique offers localization both in time and frequency [40]. Thus the DWT exhibits good frequency resolution at low frequencies and good time resolution at high frequencies. Another important advantage of the DWT is its low computational cost and ease of implementation. Hence the DWT is chosen to extract features from the EMG signal. The DWT coefficients of a signal $x(n)$ can be obtained as

$$C(a,b) = \sum_{n\in\mathbb{Z}} x[n] \psi_{a,b}[n]$$  \hspace{1cm} (2.14)

where $a$ is the dilation or scale, $b$ the translation, and $\psi_{a,b}[n]$ represents the discrete wavelet which is expressed as

$$\psi_{a,b}(n) = \left(\frac{1}{\sqrt{a}}\right) \times \psi\left(\frac{n-b}{a}\right)$$  \hspace{1cm} (2.15)

For dyadic wavelet transform, $a = 2^{-j}$, $b = k \times 2^{-j}$, $\psi_{a,b}[n] = 2^{j/2} \times \psi[2^j n - k]$ with $k \in \mathbb{Z}, j \in \mathbb{N}$.

In the discrete wavelet transform (DWT), a given signal is passed through a series of highpass filters to analyze the high frequencies and it is also passed through a series of lowpass filters to analyze the low frequencies. Here filters of different cutoff frequencies are used to analyze the signal at different scales. The procedure starts with passing the signal through a half band digital lowpass filter with impulse response $h[n]$. Filtering a signal corresponds to the mathematical operation of convolution of the signal with the impulse response of the filter given by

$$x(n) \ast h(n) = \sum_{k=-\infty}^{\infty} x(k) \cdot h(n-k).$$  \hspace{1cm} (2.16)

The DWT analyzes decomposing the signal into a coarse approximation and detail information. It employs two sets of functions, called scaling functions and wavelet functions. The original signal $x(n)$ being filtered via halfband highpass filter $h(n)$
and a lowpass filter $g(n)$ produces output of the first level decomposition, which can be respectively expressed as

$$y_{\text{high}}(k) = \sum_n x(n) \cdot h(2k - n)$$

(2.17)

and

$$y_{\text{low}}(k) = \sum_n x(n) \cdot g(2k - n).$$

(2.18)

It is to be mentioned that $y_{\text{high}}(k)$ and $y_{\text{low}}(k)$, outputs of the highpass and lowpass filters, respectively, are obtained after performing down-sampling by 2 operation. The above procedure can be repeated for further decomposition. At every level of decomposition, the filtering and down-sampling will result in half the number of samples (and hence half the time resolution) and half the frequency band spanned (and hence double the frequency resolution). In Fig. 2.3, with the help of a schematic diagram the wavelet decomposition procedure is shown for an original signal $x(n)$. In this figure, $g[n]$ and $h[n]$ are low-pass and high-pass filters, respectively. One

![Fig. 2.3: Sub-band decomposition of DWT implementation; $h[n]$ is the high pass filter, $g[n]$ the low pass filter](image)

major advantage of the DWT over DFT is that the time localization of some desired frequencies will not be lost. It means that the frequencies that are most prominent in the original signal will appear as high amplitudes in a region of the discrete wavelet transformed signal where it includes those particular frequencies. However, the time
localization will have a resolution that depends on which level they appear. The DWT offers a good time resolution at high frequencies and good frequency resolution at low frequencies. It is found that these characteristics of the DWT facilitate the analysis of the EMG signal.

The extracted wavelet coefficients provide a compact representation that shows the energy distribution of the EMG signal in time and frequency. In order to decrease the dimensionality of the extracted feature vectors, statistics over the set of the wavelet coefficients were used.

From the EMG data of a particular person, the DWT is performed on frame by frame basis. Utilizing all the DWT coefficients obtained in a frame and considering a number of such frames would make the feature dimension extremely high. Hence, from each frame, the DWT coefficients with higher values are proposed to be utilized in this research. Arranging the DWT coefficients in descending order, the first $M$ coefficients are used as features. This may still provide a large feature dimension when several frames are taken in consideration. In view of reducing the feature dimension, from each frame instead of considering $M$ coefficients, the average value of these $M$ coefficients and the maximum value of the coefficients are proposed to use as features [41].

For the purpose of comparison with the proposed wavelet domain feature, some existing proposed features in [30], are investigated.

### 2.2 Classifier: k-Nearest Neighbor (KNN)

The k-nearest neighbor (KNN) is one of the most simple but efficient classifiers used in pattern classification. It considers a distance function which is computed between the features belonging to the EMG pattern in the test set and all the EMG pattern from both normal and diseased group in the training set. In its simplest case, i.e., when $k = 1$, only one neighbor is considered for the classification and the EMG pattern to be tested is classified with the same label as the EMG pattern in the training set with the lowest value of distance. When $k > 1$, instead of being classified by only one EMG pattern from the training set, the EMG pattern from the test set is classified based on the $k$ closer EMG patterns. That is, within the $k$ closer patterns, if the normal class is in the majority, the test pattern will be classified as
normal. There are different distances that can be applied to compare features, such as Euclidean, Bhattacharyya, and Chi-squared distance. In the proposed method, the Euclidean distance is used.

A general example illustrated in Fig. 2.4 a two class problem dealing with red and blue points. Using the KNN classifier with $k = 3$, three points that are nearer to the test point is considered. It is observed from the Fig. 2.4 that the test point is closer to two red points and only one blue point. Thus, it will be classified as belonging to the red class which is the majority class. In the KNN classifier, it is required to find the optimal k value for achieving the best classification performance. In the proposed method the value of k is varied within a specified range and corresponding classification performances are analyzed.

**2.3 Database Description and Experimental Procedures**

In order to investigate the performance of the proposed disease classification scheme, a publicly available clinical EMG database (available: http://www.emglab.net/emglab/Signals/N2001/index.html) is used. A brief introduction of data acquisition and processing for this database is discussed below.
2.3.1 Signal acquisition

A concentric needle electrode is inserted in the muscle and a surface ground electrode is placed on the limb. The cable connecting the needle to the amplifier is fixed to the muscle with a piece of tape to avoid needle movement. The patient is asked to apply a slight and constant contraction. The signal is monitored both visually and by sound. An oscilloscope picture displayed on the computer monitor with a trigger and delay line system so that MUAPs that fulfill the trigger criteria are frozen on the screen. Another monitor is used to display continuously the last half second of the EMG signal. A speaker is also connected. A characteristic crispy repetitive sound is heard when the tip of the needle is close to some muscle fibers. When the examiner finds that the signal is of a usable quality a recording of a 11.2 second signal is started. The patient kept the contraction steady through the 11.2 seconds recording lasted. For conventional quantitative EMG (MUAP analysis) undistorted recordings of MUAPs are important. The needle is then moved to another level of depth or to another insertion place. Care is taken not to record from the same motor unit more than ones and to explore the whole muscle. In this way approximately twenty signals are recorded.

A standard concentric needle electrode with a leading-off area of 0.07\(\text{mm}^2\) is used. From the needle the EMG signal is fed to a high impedance differential amplifier where it is amplified and band-pass filtered using analog filter. In the amplification stage, the acquired signals are amplified 4000 times. A band pass filter with cutoff frequencies 2 Hz and 10 kHz are used. After the amplification the EMG signal is sampled and digitized at a sampling frequency of 23437.5 Hz in 16 bit resolution. The digital EMG signal is displayed on a computer monitor.

2.3.2 Patient material and recording conditions

The database consists of three different classes of data corresponding to normal control group, a group of patients with myopathy and a group of patients with ALS disease. The control group consisted of 10 normal subjects aged 21-37 years, 4 females and 6 males. 6 out of 10 were in very good physical shape, and the remaining except one were in general good shape. None in the control group had signs or history of neuronmuscular disorders. The group with myopathy consisted of 7 patients; 2
females and 5 males aged 19-63 years. All 7 had clinical and electrophysiological signs of myopathy. The ALS group consisted of 8 patients; 4 females and 4 males aged 35-67 years. Besides clinical and electrophysiological signs compatible with ALS, 5 of them died within a few years after onset of the disorder, supporting the diagnosis of ALS. The brachial biceps and medial vastus muscles are used in this study because they are the most frequently investigated in the two patient groups.

During the recording of the EMG signals, following conventional conditions for MUAP analysis are maintained: (1) the recordings are made at low (just above threshold) voluntary and constant level of contraction, (2) visual and audio feedback are used to monitor the signal quality, (3) a standard concentric needle electrode is used, (4) the EMG signals are recorded from five places in the muscle at three levels of insertion (deep, medium, low), and (5) the high and low pass filters of the EMG amplifier were set at 2 Hz and 10 kHz [42].

In the analysis of the proposed classification scheme, both two class (normal versus ALS) and three class (normal versus ALS versus myopathy) problems are taken into consideration. In order to show the effect of variation of size of the database on classification accuracy, two sizes, namely small and large datasets are used. A summary of the above four cases is provided below:

1. **Two class small dataset**: Total 30 sets of EMG recording (15 sets of normal and 15 sets of ALS data) collected from 5 normal persons and 5 ALS patients.

2. **Two class large dataset**: Total 200 sets of EMG recording (150 sets of normal and 50 sets of ALS data) collected from 10 normal persons and 8 ALS patients.

3. **Three class small dataset**: Total 45 sets of EMG recording (15 sets of normal, 15 sets of ALS and 15 sets of myopathy data) collected from 5 normal persons, 5 ALS patients, and 5 myopathy patients.

4. **Three class large dataset**: Total 250 sets of EMG recording (150 sets of normal, 50 sets of ALS and 50 sets of myopathy data) collected from 10 normal persons, 8 ALS patients, and 7 myopathy patients.

Each set of EMG recording has total 262,134 samples collected at a rate of 23,438
samples per second. Thus, each of these single channel dataset has total time duration of 11.184 sec. A single dataset is segmented into 64 distinct frames, each consisting of 4,096 samples. In order to consider the middle steady regions, from each dataset 25 frames (from 30th frame to 55th frame) are selected out of 64 frames of both the normal persons and diseased groups for feature extraction.

2.4 Results and Analysis

In this section, performance of the proposed method in classification of neuromuscular diseases is evaluated by using the publicly available database described in the previous section. In view of demonstrating the performance of the proposed disease classification scheme, following four performance parameters are utilized:

**Specificity (Sp):** Ratio of number of correctly classified healthy subjects to number of total healthy subjects.

**Sensitivity (myopathy) (SeM):** Ratio of number of correctly classified subjects suffering from myopathy to number of total subjects suffering from myopathy.

**Sensitivity (ALS) (SeA):** Ratio of number of correctly classified subjects suffering from ALS disorder to number of total subjects suffering from ALS disorder.

**Total classification accuracy (Acc):** Ratio of number of correctly classified subjects to number of total subjects.

The most widely used leave-one-out cross validation technique is used for the testing purpose. In this case, among all datasets (both for training and testing), one dataset is taken away at a time for the purpose of testing against the remaining all datasets to be used for training the classifier. Depending on the classifier’s output value, the EMG signals are classified as normal or ALS or myopathy affected EMG signals.

As discussed in previous sections, autocorrelation and ZCR are adopted as time domain features. For frequency domain features, median frequency and spectral magnitude are taken into consideration. Finally, the proposed DWT based features are used as time-frequency domain features. For the purpose of comparison, different features other than those mentioned above are taken into consideration, such as autoregressive PSDs, statistical DWT features and wavelet packet energy based features. For the sake of uniformity, in all cases, KNN classifier is employed with a
certain range of k values and similar preprocessing is performed. It is to be men-
tonred that, although the features used in the method reported in [30] are considered
here, for fair comparison, instead of the classifier used there, KNN classifier is em-
ployed. The results obtained by using these features are listed in Table 2.1. In the
table, depending on the features used in [30], namely AR, DWT, WPT, AR+DWT,
DWT+WPT and AR+WPT, corresponding schemes are abbreviated as, S-AR, S-
DWT, S-WPT, S-AR+DWT, S-DWT+WPT and S-AR+WPT respectively. It is
already mentioned that the frames of EMG signal are selected on energy basis. For
fair comparison, same frames are considered for those schemes [30]. For S-AR, the
AR parameters of a EMG frame is estimated using Burgs algorithm’s. And the
model order of the AR method is taken as 15 and window size is chosen as 256.
The AR-PSD’s are directly used as feature. In the case of S-DWT, level-5 DWT
decomposition is carried out and the following statistical measures are taken from
detail coefficients $D_1-D_5$ and approximate coefficients $A_5$. The resulting feature
vector has 23 values.

1. Mean of the absolute values of the coefficients in each sub-band.

2. Average power of the wavelet coefficients in each sub-band.

3. Standard deviation of the coefficients in each sub-band.

4. Ratio of the absolute mean values of adjacent sub-bands.

For S-WPT, wavelet packet decomposition is carried out using $db2$ mother wavelet.
As a result of decomposition, 32 feature vectors are extracted from each signal frame
and used as feature. The other schemes are combination of these feature taking two
at a time. The classification performance are reported in the Table 2.1. For two
class small dataset, S-AR+DWT produced highest classification accuracy compared
to other schemes, which is also reported in [30]. Similarly, same feature schemes
for two and three class datasets are investigated and the classification accuracy is
reported in Table 2.2, Table 2.3 and Table 2.4. It is evident from the results that
the proposed feature do not provide consistent accuracy for all datasets.

Now, for our DWT based method, it is already mentioned that a frame by frame
analysis would be carried out for feature extraction. For each frame, arranging
the DWT coefficients in a descending order, the first ten higher valued coefficients are taken as proposed features. The maximum value of those features obtained from each frame, namely the maximum DWT feature is also considered. In DWT computation, the ‘db2’ type wavelet function is used as mother wavelet. In Figs. 2.5 and 2.6, the variation of the proposed maximum DWT features for two class small dataset and two class large dataset are shown, respectively. As expected the level of proposed DWT features corresponding to the ALS patients is much higher than that of the corresponding to normal persons. It is found that the proposed features exhibit a high within class compactness and between class separability.

Fig. 2.5: Proposed maximum DWT feature for normal persons and ALS patients (Two class small dataset)

In a similar fashion, in Figs. 2.7 and Fig. 2.8, the variation of the proposed maximum DWT features for three class small and three class large datasets is shown, respectively. It is evident that due to inclusion of myopathy datasets, the classification task becomes difficult than the two class problem. However, the proposed features exhibit significant capability of distinguishing the normal, myopathy, and ALS subjects.

The performance parameters obtained by using the proposed features separately along with that obtained by using the time, frequency and time-frequency domain features are presented in Tables 2.1, 2.2, 2.3, 2.4.
Fig. 2.6: Proposed maximum DWT feature for normal persons and ALS patients (Two class large dataset)

Fig. 2.7: Proposed maximum DWT feature for normal, myopathic and ALS patients (Three class small dataset)

It is shown that the classification performance obtained by using the proposed Wavelet domain features is far superior to that obtained by using the conventional time and frequency domain features. It is to be noted that the feature dimension
Fig. 2.8: Proposed maximum DWT feature for normal, myopathic and ALS patients (Three class large dataset)

Table 2.1: Performance of proposed classification method experimented on two class small dataset

<table>
<thead>
<tr>
<th>Feature set</th>
<th>Value of $K$</th>
<th>Sp</th>
<th>SeA</th>
<th>Acc in percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>S-AR</td>
<td>10</td>
<td>83.4667</td>
<td>79.6000</td>
<td>81.5333</td>
</tr>
<tr>
<td>S-DWT</td>
<td>8</td>
<td>77.4000</td>
<td>77.4000</td>
<td>74.4000</td>
</tr>
<tr>
<td>S-WPT</td>
<td>7</td>
<td>66.8000</td>
<td>63.2667</td>
<td>65.0333</td>
</tr>
<tr>
<td>S-AR+WPT</td>
<td>8</td>
<td>82.9333</td>
<td>79.8667</td>
<td>81.4000</td>
</tr>
<tr>
<td>S-AR+DWT</td>
<td>10</td>
<td>82.6667</td>
<td>80.0000</td>
<td>81.3333</td>
</tr>
<tr>
<td>S-DWT+WPT</td>
<td>7</td>
<td>77.0667</td>
<td>77.9333</td>
<td>77.5000</td>
</tr>
<tr>
<td>Time domain: ACR</td>
<td>3</td>
<td>100.0000</td>
<td>93.3333</td>
<td>96.6667</td>
</tr>
<tr>
<td>Time domain: ZCR</td>
<td>5</td>
<td>66.6667</td>
<td>60.0000</td>
<td>63.3333</td>
</tr>
<tr>
<td>Freq domain: FFT</td>
<td>1</td>
<td>86.6667</td>
<td>86.6667</td>
<td>86.6667</td>
</tr>
<tr>
<td>Freq domain: MDF</td>
<td>10</td>
<td>86.6667</td>
<td>40.0000</td>
<td>63.3333</td>
</tr>
<tr>
<td>DWT Five Coeff</td>
<td>1</td>
<td>100.0000</td>
<td>100.000</td>
<td>100.000</td>
</tr>
<tr>
<td>DWT Fifteen Coeff</td>
<td>7</td>
<td>100.0000</td>
<td>93.3333</td>
<td>96.6667</td>
</tr>
<tr>
<td>Proposed Method-1: DWT Max</td>
<td>1</td>
<td>100.0000</td>
<td>100.0000</td>
<td>100.0000</td>
</tr>
<tr>
<td>Proposed Method-2: DWT Ten Coeff</td>
<td>1</td>
<td>100.0000</td>
<td>100.0000</td>
<td>100.0000</td>
</tr>
</tbody>
</table>
Table 2.2: Performance of proposed classification method experimented on two class large dataset

<table>
<thead>
<tr>
<th>Feature set</th>
<th>Value of K</th>
<th>Sp</th>
<th>SeA</th>
<th>Acc in percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>S-AR</td>
<td>7</td>
<td>91.7800</td>
<td>48.9400</td>
<td>81.0700</td>
</tr>
<tr>
<td>S-DWT</td>
<td>10</td>
<td>92.0000</td>
<td>44.8400</td>
<td>80.2100</td>
</tr>
<tr>
<td>S-WPT</td>
<td>7</td>
<td>90.4467</td>
<td>32.9800</td>
<td>76.0800</td>
</tr>
<tr>
<td>S-AR+WPT</td>
<td>7</td>
<td>91.7867</td>
<td>48.9600</td>
<td>81.0800</td>
</tr>
<tr>
<td>S-AR+DWT</td>
<td>10</td>
<td>93.5800</td>
<td>44.7800</td>
<td>81.3800</td>
</tr>
<tr>
<td>S-DWT+WPT</td>
<td>7</td>
<td>77.0667</td>
<td>77.9333</td>
<td>77.5000</td>
</tr>
<tr>
<td>Time domain:ACR</td>
<td>7</td>
<td>91.1533</td>
<td>45.3000</td>
<td>79.6900</td>
</tr>
<tr>
<td>Time domain:ZCR</td>
<td>5</td>
<td>92.0000</td>
<td>64.0000</td>
<td>85.0000</td>
</tr>
<tr>
<td>Freq domain:FFT</td>
<td>7</td>
<td>94.0000</td>
<td>34.0000</td>
<td>79.0000</td>
</tr>
<tr>
<td>Freq domain:MDF</td>
<td>1</td>
<td>83.3333</td>
<td>30.0000</td>
<td>70.0000</td>
</tr>
<tr>
<td>DWT Five Coeff</td>
<td>4</td>
<td>92.6667</td>
<td>72.0000</td>
<td>87.5000</td>
</tr>
<tr>
<td>DWT Fifteen Coeff</td>
<td>1</td>
<td>92.6667</td>
<td>72.0000</td>
<td>85.5000</td>
</tr>
<tr>
<td>Proposed Method-1:</td>
<td>11</td>
<td>96.6667</td>
<td>74.0000</td>
<td>91.0000</td>
</tr>
<tr>
<td>DWT Max</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proposed Method-2:</td>
<td>8</td>
<td>97.3333</td>
<td>74.0000</td>
<td>91.5000</td>
</tr>
<tr>
<td>DWT Ten Coeff</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

is drastically reduced when the maximum wavelet feature is used. However, these two features still provide an excellent recognition performance. So, the proposed method is effective which gives highest success rate in distinguishing the EMG data of neuromuscular patients from the normal control group.

2.5 Conclusion

In this chapter, first, a preprocessing technique is presented, which is important prior to feature extraction from raw EMG signal. In the proposed method, apart from the conventional time and frequency domain features, some DWT based features are introduced to handle both the two classes and three classes problem of separating the EMG signals of normal, myopathic and ALS patients. Different sizes of databases are formed to validate the classification performance. It is found that the DWT based proposed features extracted from EMG data can exhibit high within class compactness and between class separability. Apart from the classification performance, the proposed features exhibits reduced computational complexity due to
Table 2.3: Performance of proposed classification method experimented on three class small dataset

<table>
<thead>
<tr>
<th>Feature set</th>
<th>Value of $K$</th>
<th>$Sp$</th>
<th>$SeA$</th>
<th>$SeM$</th>
<th>Acc in percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>S-AR</td>
<td>10</td>
<td>68.1333</td>
<td>76.0667</td>
<td>81.5333</td>
<td>75.2444</td>
</tr>
<tr>
<td>S-DWT</td>
<td>7</td>
<td>56.4000</td>
<td>75.2000</td>
<td>76.7333</td>
<td>69.4444</td>
</tr>
<tr>
<td>S-WPT</td>
<td>1</td>
<td>45.6000</td>
<td>56.0667</td>
<td>60.3333</td>
<td>54.0000</td>
</tr>
<tr>
<td>S-AR+WPT</td>
<td>7</td>
<td>68.5333</td>
<td>74.0667</td>
<td>80.6000</td>
<td>74.4000</td>
</tr>
<tr>
<td>S-AR+DWT</td>
<td>7</td>
<td>66.2000</td>
<td>74.8667</td>
<td>81.0667</td>
<td>74.0444</td>
</tr>
<tr>
<td>S-DWT+WPT</td>
<td>7</td>
<td>53.4667</td>
<td>69.6000</td>
<td>70.0667</td>
<td>64.3778</td>
</tr>
<tr>
<td>Time domain:ACR</td>
<td>3</td>
<td>73.3333</td>
<td>93.3333</td>
<td>53.3333</td>
<td>73.3333</td>
</tr>
<tr>
<td>Time domain:ZCR</td>
<td>7</td>
<td>46.6667</td>
<td>60.0000</td>
<td>53.3333</td>
<td>53.3333</td>
</tr>
<tr>
<td>Freq domain:FFT</td>
<td>1</td>
<td>80.0000</td>
<td>62.0000</td>
<td>20.0000</td>
<td>71.6279</td>
</tr>
<tr>
<td>Freq domain:MDF</td>
<td>1</td>
<td>84.0000</td>
<td>44.0000</td>
<td>26.6667</td>
<td>70.6977</td>
</tr>
<tr>
<td>DWT Five Coeff</td>
<td>4</td>
<td>80.0000</td>
<td>100.0000</td>
<td>80.0000</td>
<td>86.6667</td>
</tr>
<tr>
<td>DWT Fifteen Coeff</td>
<td>9</td>
<td>73.3333</td>
<td>93.3333</td>
<td>86.6667</td>
<td>84.4444</td>
</tr>
<tr>
<td>Proposed Method-1:DWT Max</td>
<td>7</td>
<td><strong>73.3333</strong></td>
<td><strong>100.0000</strong></td>
<td><strong>93.3333</strong></td>
<td><strong>88.8889</strong></td>
</tr>
<tr>
<td>Proposed Method-2:DWT Ten Coeff</td>
<td>3</td>
<td><strong>86.6667</strong></td>
<td><strong>100.0000</strong></td>
<td><strong>86.6667</strong></td>
<td><strong>91.1111</strong></td>
</tr>
</tbody>
</table>

good quality feature and small feature dimension. Disease classification performance is evaluated on one publicly available standard clinical EMG database and it is observed that the proposed scheme provides satisfactory performance in comparison to some of the existing methods both in terms of classification accuracy as well as computational complexity.
Table 2.4: Performance of proposed classification method experimented on three class large dataset

<table>
<thead>
<tr>
<th>Feature set</th>
<th>Value of K</th>
<th>Sp</th>
<th>SeA</th>
<th>SeM</th>
<th>Acc in percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>S-AR</td>
<td>7</td>
<td>82.7467</td>
<td>48.1600</td>
<td>57.8000</td>
<td>70.8400</td>
</tr>
<tr>
<td>S-DWT</td>
<td>7</td>
<td>80.8800</td>
<td>46.7600</td>
<td>52.0800</td>
<td>68.2960</td>
</tr>
<tr>
<td>S-WPT</td>
<td>7</td>
<td>77.3200</td>
<td>33.3400</td>
<td>32.2800</td>
<td>59.5160</td>
</tr>
<tr>
<td>S-AR+WPT</td>
<td>1</td>
<td>82.7533</td>
<td>48.2200</td>
<td>57.8400</td>
<td>70.8640</td>
</tr>
<tr>
<td>S-AR+DWT</td>
<td>1</td>
<td>82.1800</td>
<td>46.9800</td>
<td>59.2000</td>
<td>70.5440</td>
</tr>
<tr>
<td>S-DWT+WPT</td>
<td>1</td>
<td>80.8933</td>
<td>46.9000</td>
<td>51.9600</td>
<td>68.3080</td>
</tr>
<tr>
<td>Time domain:ACR</td>
<td>7</td>
<td>84.6667</td>
<td>64.0000</td>
<td>52.0000</td>
<td>74.0000</td>
</tr>
<tr>
<td>Time domain:ZCR</td>
<td>5</td>
<td>82.6667</td>
<td>30.0000</td>
<td>22.0000</td>
<td>60.0000</td>
</tr>
<tr>
<td>Freq domain:FFT</td>
<td>7</td>
<td>84.0000</td>
<td>28.0000</td>
<td>38.0000</td>
<td>63.6000</td>
</tr>
<tr>
<td>Freq domain:MDF</td>
<td>1</td>
<td>76.6667</td>
<td>28.0000</td>
<td>48.0000</td>
<td>61.2000</td>
</tr>
<tr>
<td>DWT Five Coeff</td>
<td>6</td>
<td>89.3333</td>
<td>70.0000</td>
<td>90.0000</td>
<td>85.6000</td>
</tr>
<tr>
<td>DWT Fifteen Coeff</td>
<td>3</td>
<td>89.3333</td>
<td>72.0000</td>
<td>84.0000</td>
<td>84.8000</td>
</tr>
<tr>
<td>Proposed Method-1:DWT Max</td>
<td>17</td>
<td>92.0000</td>
<td>72.0000</td>
<td>82.0000</td>
<td>86.0000</td>
</tr>
<tr>
<td>Proposed Method-2:DWT Ten Coeff</td>
<td>8</td>
<td>91.3333</td>
<td>74.0000</td>
<td>90.0000</td>
<td>87.6000</td>
</tr>
</tbody>
</table>
Chapter 3

Proposed MUAP Signal Based Disease Classification Scheme

Electromyography plays a major role in physiological investigations and clinical examinations for either the study of motor control or the diagnosis of neuromuscular disorders. There are two major techniques of such investigations: Raw EMG signal and EMG signal decomposition. In the previous chapter, it is presented that a frame by frame raw EMG signal analysis provides good classification accuracy for neuromuscular disease classification. It is to be noted that if a single frame is to be classified, the frame should contain sufficient underlying information for consistent performance. However, our proposed method takes continuous frames from each EMG dataset in order to extract global statistics rather than taking single frame, which automatically resolves the consistency problem. In the other way around, EMG signal decomposition can be another approach to analyze the signal. EMG signal decomposition provides a unique way to observe the behavior of individual motor neurons in the intact human nervous system. Also, since the shapes of the motor unit action potential (MUAP) convey information about the characteristics and arrangement of the muscle fibers, EMG decomposition providing a unique way to study motor-unit organization in intact human muscles. This information can be a great use in clinical neurophysiology for diagnosing neuromuscular disorders.

In this chapter, an efficient scheme for neuromuscular disease classification based on MUAP is proposed with two fold objectives: (1) to define and select dominant MUAP that is present in the EMG signal bearing maximum information, based on temporal energy and (2) to extract wavelet based feature from dominant MUAPs for classification and taking decisions. First, the existing MUAP based analyses are
presented wherein the methods for extracting MUAPs from EMG signal are of two categories: 1) Direct MUAP collection from needle electrode and 2) Decomposition of EMG recording. Our proposed method used the later approach using a publicly available template matching based EMG decomposition tool (EMGLAB) to acquire MUAP from EMG signal. Conventional feature extraction technique from MUAP along with the proposed analysis is discussed in the following sections. The morphological feature such as amplitude, duration, rise time, area, and no. of phases of MUAP are considered as conventional feature. Then the proposed features from spectral domain (DCT) and wavelet domain (DWT) are obtained. The classification performance is evaluated on the same database discussed in the previous chapter. Our proposed DWT based statistical feature provides excellent classification accuracy of 100% in small database and of 98.8% in large database. Finally, all of the results and analysis are reported in the following section.

3.1 MUAP Recording and Characteristics

The interpretation of the electrical activity of muscle is subjective to some factors that require quantification for exact evaluation. The diagnoses of neuromuscular disorders are dependent to these electrophysiologic factors discussed as follows.

3.1.1 Motor Unit (MU)

The motor unit (MU) is the smallest functional unit of a skeletal muscle, and its electrophysiologic characteristics are central in the diagnosis of neuromuscular disorders associated with partial denervation and myopathy. It comprises the alpha motor neurone, its axon, the motor end-plates and all muscle fibres innervated by this axon. The number of MUs in different muscles vary considerably and the exact numbers are not known. Neither is there exact information about number of muscle fibres per MU. All muscle fibres of a given MU have the same histochemical and biomechanical characteristics. The MU size, expressed as number of fibres or anatomical cross-sectional area, varies considerably between MUs within the same muscle. In normal MUs muscle fibres are randomly scattered within an area called the MU territory. In diseases muscle fibre topography within the MU territory and the interstitial tissue change. This produces characteristic changes in the EMG
signals since there is a relatively close relationship between the MU structure and electrophysiological events. EMG can therefore be used to assess structural alterations of MUs in individual muscles and the distribution of abnormalities in the limb muscles (proximal, distal or other distributions).

The electrical activity of the MU can be recorded in different ways. Depending upon the technique, smaller or larger portions of the motor unit can be studied. The monopolar electrode with a remote reference electrode, the concentric needle electrode recording is one of the recording scheme. In SFEMG, with a very small recording surface, a selective recording is obtained for the detailed study of individual muscle fibres and motor end-plates within a motor unit. On the other hand, with Macro EMG, a nonselective recording of the total electrical size (comprising the number and size of individual muscle fibres in a motor unit) may be obtained. Surface EMG records activity from a large part of a muscle or even many neighbouring muscles, reflecting the electrical activity of muscles within a large area [42].

3.1.2 Motor Unit Action Potential (MUAP)

The MUAP is a compound signal reflecting the summation and cancellation of phases of the action potentials from individual muscle fibers in the motor unit. With intracellular recordings, the action potential is a monophasic waveform of about 100mV, whereas the extracellular potential is a volume-conducted derivative of the rate of membrane depolarization. The MUAP represents the spatial and temporal summation of these bi(tri)phasic spikes, wherein the negative spike in the normal muscle is obtained from two to three fibers within 0.5 to 1mm of the electrode. The amplitude of the spike is determined by the proximity of the closest active fibers as indicated by the fact that fibrillation potentials that originate from single fibers may have as high an amplitude as the MUAP.

The shape and duration of the MUAP reflects the architecture of the motor unit. Recorded outside the end-plate region, the MUAP typically has three phases: an initial positive phase, a negative spike, and a terminal positive phase. In some instances, the MUAP may be split up into four or more phases (polyphasic), reflecting a greater asynchrony of individual muscle fiber discharges.

A MUAP may contain a spike component separate from the main spike, a so-
called satellite potentials which represent action potentials of a single or a few fibers temporally dispersed from the main bulk of the fibers in the motor unit in disease states and contributes up to 3% of MUAPs of healthy muscle [42].

3.1.3 Morphological Parameters of MUAP

The major morphological MUAP parameters are of clinical importance: the amplitude, duration, rise time, area and the number of phases. A sample MUAP is shown in Fig. 3.1 for defining the parameter.

Fig. 3.1: Illustration of the MUAP parameters

Amplitude of the Motor Unit Action Potential

MUAP amplitude is measured peak-to-peak from the most positive to the sequentially most negative peak. Because of cancelation between phases, the MUAP amplitude is less than the sum of individual fiber potentials and may even be smaller than the amplitude of single fiber potentials. It depends on the proximity of the closest 2 to 15 fibers of the motor unit within about a $0.5\, \text{mm}$ diameter and is proportional to the number and density of fibers in the motor unit. The amplitude also depends on the type of electrode used. The amplitude is larger when using monopolar than
concentric needle electrodes. It is important to hear the crisp sound of discharging MUAPs indicating that the electrode is placed close to active motor units. The amplitude tends to increase in chronic partial denervation with reinnervation of muscle fibers, whereas it is shorter in myopathy due to loss of muscle fibers.

**Duration of the Motor Unit Action Potential**

The duration of the MUAP reflects the temporal dispersion of activity of fibers constituting the motor unit and is primarily due to the spatial distribution of end-plates along muscle fibers measuring 20 to 30mm, with a conduction velocity of 3 to 5m/sec [19].

A duration feature defined in the literature is a bit different. It is the time interval between MUAP onset and offset points. To define the onset of the MUAP waveform, the first point is identified where the signal is greater than a threshold equal to 1/15 of the amplitude. Starting from this point and moving backward to the beginning of the waveform, a sliding window of 1ms is applied. The point in the window closer to the baseline is the MUAP onset point. The MUAP offset point is calculated in a similar way.

MUAP duration is one of the most important parameter in the separation of myopathic disorders in which there is a loss of muscle fibers from chronic neurogenic disorders in which there is an increased number of muscle fibers due to collateral sprouting [42].

**Rise Time**

It is the time between the maximum negative peak and subsequent maximum positive peak within the duration of each MUAP.

**Area**

It is defined between the MUAP waveform and the baseline for the calculated duration.

**Number of phases**

A phase is a section of a MUAP that falls between two baseline crossings and reaches an absolute value of amplitude larger than 0.02mV. They are counted within each
MUAP. MUAPs can be simple in shape or polyphasic. The latter is of long or short duration in myopathy depending on the degree of muscle fiber regeneration. It is important to calculate amplitudes and durations of simple and polyphasic potentials separately and to collect more MUAPs than necessary in case additional ones are needed. Irregularities of the MUAP that do not result in baseline crossings are termed “turns” if they have an amplitude of more than 100\(\mu\)V (Fig. 3.1). Studies of MUAPs have shown variability in the shape in successive discharges or jiggle [21] not due to electrode displacement. This is more pronounced in myasthenia gravis and in early collateral sprouting with immature axonal sprouts.

**Significance of MUAP in Disease Classification**

The primary aim of the EMG examination is to determine whether weakness is due to a myopathic or neurogenic lesion. Neuromuscular disorders change the morphology and physiology of MUs causing changes in their activation patterns and MUAP shapes and thus the EMG signals that they produce. Myopathic disorders occur when muscle fibers die or atrophy, e.g. Muscular Dystrophy. Neuropathic disorders occur when motor neurons die and the remaining surviving motor neurons reinnervate orphaned muscle fibers, e.g. Amyotrophic Lateral Sclerosis (Lou Gehrig’s disease). Fig. 3.2 shows that in general the size and shape of MUAPs detected from muscles affected by a myopathic disorder are smaller and more complex while the shapes of MUAPs detected from neuropathic muscles are larger and more complex compared to MUAPs detected from normal muscles. However, in practice there is a great deal of ambiguity in the interpretation of MUAPs detected from a muscle as MUAPs that appear normal, myopathic and neuropathic can be detected from any muscle regardless of its condition [43]. This ambiguity exists because of how widely muscle structure and needle position can vary.

The basic pathophysiologic change in myopathy is the degeneration loss of muscle fibers in the individual motor unit. Whether these abnormalities are reflected in the EMG depends on the stage of the disease and whether compensatory regeneration mechanisms have already occurred. At early stages of the disorder, muscle contraction may be affected primarily at the level of excitation-contraction coupling, in which case MUAP analysis may show only mild or unspecific changes. The incidence of polyphasic potentials is increased in patients with myopathy for two reasons.
First, muscle fibers of the motor units are lost and hence summation of muscle fiber action potentials in the MUAP is reduced. Because of the loss of muscle fibers, the duration of the MUAP is reduced (Fig. 3.3). Second, there is regeneration of muscle fibers from satellite cells, and this is associated with polyphasic potentials due to increased dispersion caused by a reduced conduction velocity along the regenerated
Fig. 3.3: MUAPs from a normal subject, a patient with neuropathy and a patient with myopathy

muscle fibers or along collateral nerve sprouts.

In neurogenic lesions, the interference pattern is reduced in moderate cases with loss of motor units, whereas it is discrete in severe cases wherein individual MUAPs stand out with a flat baseline between successive discharges. In chronic motor neuron disorders with reinnervation due to collateral sprouting, the amplitude of the reduced or discrete interference pattern is increased and may reach values of 8 to 10 mV. The distribution, type, and degree of EMG abnormalities are important in the diagnosis of neuromuscular disorders. Although the EMG examination should include clinically affected muscles, it may be crucial to study muscles not overtly affected. For example, in patients with suspected motor neuron disease, the initial weakness and atrophy may have a segmental distribution, raising suspicion of either radicular affection or mononeuropathy. However, EMG examination of muscles with normal force and bulk often shows that the involvement is distributed more widely than suggested by the clinical examination and thus indicate a generalized disorder [42].

3.2 EMG Decomposition for MUAP Extraction

The purpose of EMG signal decomposition is to separate the composite interference pattern into its constituent motor unit action potential (MUAP) shapes, firing times, permitting the evaluation and study of individual MU firing patterns. As
discussed earlier, MUAP can be obtained either by precise needle insertion into the muscle fiber or decomposition of raw EMG signal. Both of these processes have their own pros and cons. However, the decomposition process has some advantages over the needle MUAP extraction. The needle MUAP can easily be interfered by nearby neighbor muscle fiber, resulting in interference pattern (IP). On the other hand, by means of decomposition, this IP pattern will break down to its constituent parts (i.e. MUAP). Hence, it is feasible to decompose the raw EMG signal for obtaining MUAPs. There are numerous techniques available for EMG decomposition. Recently, Kevin McGill [8] has developed an open source graphical EMG decomposition tool for single-channel EMG recording. Some other decomposition techniques are proposed in [20], [44], [10] and [11]. The objective of thesis is to find dominant MUAPs from each EMG signal. These dominant MUAPs are the MUAP bearing highest temporal energy content. EMGLAB decomposes the signal by “template matching” algorithm. First, it reads the initial 2 sec of the EMG signal and automatically creates templates for all the spikes that occur at least three times with a high-degree of similarity. Then, the program tries to automatically classify the remaining spikes in this 2 sec interval using template matching. It is also able to resolve many superimpositions during the automatic template-matching phase by identifying larger spikes and “peeling them off” to reveal smaller spikes in the residual. It is observed experimentally that EMGLAB does not usually miss the high energy content MUAPs, which is our prior requirement. EMGLAB also uses an averaging algorithm to subtract out the effect of the interference from the other MUAPs for higher signal-to-noise ratios, which is another advantage of this tool. A snapshot of EMGLAB decomposition tool and extracted MUAPs from a 11.2 sec EMG signal are shown in Fig. 3.4 and Fig. 3.5 respectively. The EMGLAB facilitates the MUAP extraction process quite efficiently, hence, employed for MUAP extraction. More background and procedure of decomposition is given in Appendix A.
3.3 Proposed Dominant MUAP Selection Method

3.3.1 Relation between MUAP energy and motor unit characteristics

A motor unit is generally affected by neuromuscular disorders, such as muscle fibre loss in case of myopathy or muscle fibre gain in case of ALS through collateral sprouting in response to nearby denervation. Both pathological situations affect motor unit size (spatial extent and muscle fibre number) and muscle fibre density. The physiological effect of muscle fibre loss is to weaken the force-generating capacity of the motor unit and the muscle. In contrast, collateral sprouting is a compensatory mechanism designed to mitigate the effect of denervation, so as to maintain overall muscle strength by increasing the force-generating capacity of participating motor units.

Since the MUAP of a certain motor unit can be considered as a summation of fibre potentials of all muscle fibres belonging to that motor unit, above pathophys-
Fig. 3.5: Template panel, showing the templates of the identified MUAPs

iological changes will affect the amplitude of MUAPs in disease groups [42]. For example, in case of myopathy, MUAPs become low in amplitude and short in duration. In order to compensate for the low amplitude, a larger number of motor units are recruited at lower than normal levels of muscle contraction. Full interference pattern is developed at less than maximal contraction with increasing force. On the other hand, in neurogenic disorders (ALS), the excited motor neurons are decreased in number and muscle fibres are reorganized to give larger motor units, this causes MUAPs having higher amplitude and longer duration than normal. In order to keep a certain force of contraction, the available motor neurons must fire at a higher rate than normal to balance the motor neuron loss [30]. There is potentially an attractive link between motor unit size and force-generation and the energy content of a MUAP. The concept of a MUAP having energy is readily intuitive when one recalls the deep booming sound of a neurogenic MUAP that sometimes rattles the EMG machine [45]. So it is evident that the energy of individual MUAP will be different in different groups (i.e. normal, ALS, and myopathy). Hence, the energy content of MUAPs provide significant information about the EMG signal and idea about pathology.
3.3.2 Energy content based dominant MUAP selection

In MUAP based disease classification methods, generally all of the extracted MUAPs are equally treated, no matter whether the MUAPs are extracted by direct acquisition from the muscle or via decomposition of EMG signal. In both types of MUAP extraction process, a good number candidate MUAPs are obtained. However, it is well known that not all of them can uniquely characterize the class they belong to. Possible reasons behind such discrimination could be the variation in firing characteristics, cross talk (interference from neighboring muscle fibers), and acquisition process. The firing rate of some MUAPs are very less than the normal rate as well. Hence, we propose to use only the dominant MUAP among the MUAPs extracted from a particular EMG recording for the purpose of classification.

As mentioned above that the energy content of MUAPs plays an important role in discriminating different types of subjects, the selecting criteria for dominant MUAP is proposed to be the temporal energy content of the MUAP. The energy of a signal $x(n)$ is given by

$$E_x = \sum_{0}^{N-1} x(n)^2.$$  \hspace{1cm} (3.1)

It is to be mentioned that the energy content of MUAP is generally higher for ALS group and and lower for myopathy group with respect to the energy content of MUAP of normal group. Among the extracted MUAPs from a particular EMG recording, the MUAP with the highest energy content is selected as the dominant MUAP. Using only the dominant MUAP instead of all MUAPs during training and testing phases will not only offer significant reduction in computational complexity but also provide better consistency in feature quality.

It is to be noted that different number of MUAPs are expected from EMG decomposition. Generally, the number varies from 3 – 18. However, the dominant MUAPs are selected and considered for further analysis. Temporal patterns of all of the acquired MUAPs extracted via the decomposition of a normal EMG recording is shown in Fig. 3.6. The energy contents corresponding to the MUAPs shown in Fig. 3.6 are plotted in Fig. 3.7. It is observed that the second(i.e. No. B) MUAP contains the highest temporal energy and hence identified as the dominant MUAP for that particular EMG recording. In a similar fashion, all decomposed MUAPs
Fig. 3.6: MUAPs after decomposition of single normal EMG signal

from one set of EMG recordings of ALS and myopathy group are shown in Fig. 3.8 and Fig. 3.10. And the corresponding identified dominant MUAPs are shown in Fig. 3.9 and Fig. 3.11 respectively. As expected the energy content of the ALS group is the highest followed by the normal group and the myopathic dominant MUAP has the lowest energy. It is to be noted that for selecting the dominant MUAP, the spectral energy content is also investigated and similar results are obtained. Once the dominant MUAPs for different datasets are obtained, these are then used for the feature extraction.

Firing rate of the MUAP (the number of occurrence of a particular MUAP in a EMG recording) is also inquired for selection of the dominant MUAP. Quantifying the changes in firing rate and variability from the firing patterns can reveal changes in the central nervous system and can help in differentiating neuromuscular disorders compared to normal condition. But one prior condition here is to find the exact firing pattern, which can only be obtained by full EMG decomposition. The selection of dominant MUAP considering the highest firing rate may not be very suitable due to its complete dependence on the decomposition process.
Feature extraction is a crucial step for EMG based neuromuscular disease classification, which directly dictates the classification accuracy. With a view to obtain the spectral characteristics of MUPs, although discrete Fourier transform (DFT) can be employed, in this work, we propose to employ the discrete cosine transform (DCT) based feature extraction scheme. The DCT of a signal \( x(n) \) is computed as

\[
y(k) = w(k) \sum_{n=1}^{N} x(n) \cos \frac{\pi(2n-1)(k-1)}{2N}, \quad k = 1, \ldots, N,
\]

where, \( w(k) = \begin{cases} \frac{1}{\sqrt{N}}, & k = 1 \\ \frac{\sqrt{2}}{\sqrt{N}}, & 2 \leq k \leq N \end{cases} \)

The reason behind choosing DCT is that it is superior to DFT for the transformation of real signals. For a real signal, the DFT gives a complex spectrum and leaves nearly half of the data unused. In contrast, the DCT generates a real spectrum of real signals and thereby avoids the computation of redundant data. Moreover, it offers an ease of implementation in practical applications. The energy compaction property of DCT allows representation in lower dimensions. This facilitates reducing the number of coefficients to be employed as feature in an intended classification task. Due to the strong energy compaction property of DCT, since most of the important information tends to be concentrated in a few low frequency DCT coefficients, better noise immunity is expected.
The DCT of dominant MUAPs corresponding to normal, ALS, and myopathy subjects are investigated. It is found that they are quite dissimilar from each other and definitely can be a distinguishable feature.

For a single dataset of normal group, the absolute DCT coefficients of dominant MUAP are shown in Fig. 3.12. It is observed from the figure that high energy coefficients are mainly present in low frequency region. Similarly, the absolute DCT coefficients of the dominant MUAP corresponding to ALS and myopathy patients are shown in Figs. 3.13 and 3.14 respectively. As expected, the maximum value in case of ALS patient is greater than that of normal person or myopathy patient and the maximum value in case of myopathy patient group is found to be the lowest among all three cases. Hence, maximum value of absolute DCT coefficients of MUAP could be a potential feature.

Apart from the maximum value, the mean value of absolute DCT coefficients of MUAP corresponding to a particular dataset is also investigated. Considering the large dataset in Fig. 3.15, the mean values of absolute DCT coefficients of dominant MUAPs obtained from different EMG recordings corresponding to normal, ALS and myopathy subjects are shown. It is evident from the figure that the mean values for the three different classes are very much overlapping. Thus, mean value of
absolute DCT coefficients of MUAP may not be considered as a potential feature for neuromuscular disease classification.

In a similar fashion, considering the same large dataset in Fig. 3.16, the maximum values of absolute DCT coefficients of dominant MUAPs obtained from different EMG recordings corresponding to normal, ALS and myopathy subjects are
Fig. 3.11: Energy content of extracted MUAPs from single myopathy EMG signal shown. It is observed from the figure that the maximum values for the three different classes are well separated. Hence, maximum value of absolute DCT coefficients of MUAP is considered as the proposed feature for neuromuscular disease classification.

Fig. 3.12: Absolute DCT coefficients of dominant MUAP extracted from a normal person
3.5 Proposed DWT Based Feature

Wavelet analysis was carried out on dominant MUAP’s recorded from normal subjects and subjects suffering with motor neuron disease and myopathy. The frequency content of MUAP signal provides useful information than time domain representation. The wavelet transform gives the multi-resolution description of a nonstationary signal. Since EMG is non stationary signal hence wavelet is suited for EMG signal
Fig. 3.15: Mean values of absolute DCT coefficients of dominant MUAPs obtained from different EMG recordings corresponding to normal, ALS and myopathy subjects

Fig. 3.16: Maximum values of absolute DCT coefficients of dominant MUAPs obtained from different EMG recordings corresponding to normal, ALS and myopathy subjects

analysis, so as the MUAP. At high frequencies it represents a good time resolution and for low frequencies it represents better frequency resolution. This multi-scale feature of the Wavelet allows the decomposition of a signal into a number of scales, each scale representing a particular coarseness of the signal under study. The procedure of multiresolution decomposition of a signal $x[n]$ is schematically shown in
chapter 2. The approximation coefficients roughly represent the magnitude of the signal at that time point, which corresponds to the peak of the wavelet function. And the detail coefficients represent successively higher frequency information that is absent from the approximation.

In the proposed method, the DWT analysis is carried out on dominant MUAP. In Fig. 3.17, the extracted approximate coefficients (A1) and detail coefficients (D1) from a normal dominant MUAP is shown. Similarly, an ALS affected and myopathy affected dominant MUAP and its level-1 DWT decomposition is shown in Fig. 3.18 and Fig. 3.19 respectively. The extracted wavelet coefficients provide a compact representation that shows the energy distribution of the MUAP signal in time and frequency band. The wavelet coefficients are computed using daubechies wavelet of order 2 because its smoothing features are more suitable to detect changes in MUAP signal. In [18], the DWT analysis is carried out on all MUAP. They utilized level-6 DWT decomposition and extracted feature from all details coefficient and latest approximate coefficients. Furthermore, they proposed to use some heuristically selected coefficients that is not viable for all MUAPs. Since, detail coefficients mainly concentrates on higher frequency information, it is well known that MUAP does not attribute information in higher frequency region. In the contrast, most of the significant information are found in a low range of frequency. A multiple level of DWT decomposition is carried out on normal EMG signal and shown in Fig. 3.20. It is seen from the Fig. 3.20 that upto level-3 decomposition the approximate waveforms are quite similar and it is not feasible to extract information from them. However, level-4 decomposition waveform gives information about peaks on lower frequency region. In addition to that, the proposed work is dependent on the length of MUAP signal so that level-6 DWT decomposition can be carried out. In view of these limitation, we propose to consider level-1 DWT decomposition and investigated both the approximate and details coefficients.

Therefore, the computed detail and approximation wavelet coefficients of the dominant MUAP signals are utilized as the feature vectors representing the signal. In order to reduce the dimensionality of the feature vectors, statistics over the set of the wavelet coefficients are employed. The following statistical features are used to represent the time-frequency distribution of the dominant MUAP signals:
Fig. 3.17: Wavelet decomposition of a normal dominant MUAP

1. Maximum of the approximate wavelet coefficients.

2. Standard deviation of the approximate wavelet coefficients.


In [18], the DWT coefficients are taken from level-6 of detail coefficients. Since, the MUAP does not inevitably contain much information in low frequency, level-6 decomposition of all MUAP may not exhibit the same accuracy. It is evident from our proposed method that level-1 decomposition is sufficient to get the desire feature for separating EMG signals.

As per [18], wavelet coefficients with higher energy concentration from both the approximate and details coefficients can be considered as feature. Therefore, in the proposed feature vector, the maxima value of both the approximate and detail coefficients are considered. The standard deviation of wavelet coefficients are also
included in the feature vector. Alongwith this feature, mean and standard deviation statistical measures are considered. Mean of wavelet coefficients is first considered as feature. It is apparent that the mean value feature of dominant MUAPs of different class overlaps and exhibits degraded performance. Hence mean feature of coefficients are discarded. Apart from mean feature, standard deviation of coefficients performs quite well. Finally the feature vector are formed with maxima and standard deviation values of approximate and detail coefficients.

These feature vectors are then given input to the classifier. The performance of wavelet analysis is also compared to traditional frequency domain analysis. Consequently, the maximum magnitude of discrete cosine transform (DCT), spectral amplitude and mean frequency features are investigated.
3.6 Results and Analysis

The decomposition of raw EMG signals into their constituent MUAPs and their classification into groups is a typical classification problem. A robust and reliable method is yet to propose by the researchers. In this thesis, one such method is proposed with its validation using a publicly available clinical raw EMG database. Our proposed algorithm of disease classification is evaluated in terms of specificity ($Sp$), sensitivity ALS ($SeA$), sensitivity myopathy ($SeM$), and total classification accuracy $TAcc$ defined in section: described in the previous chapter. The disease classification scheme based on MUAP signal proposed in this chapter consists of two stages: (1) Selection of dominant MUAP from all decomposed MUAPs for each EMG dataset. (2) Classification of dominant MUAP representing the EMG signal by quality feature.

The database arrangements are same as described in the previous chapter. First, the raw EMG signal is decomposed by EMGLAB for acquiring its constituent
MUAPs. The auto decomposition feature is carried on the 11.2 sec EMG signal in three 5 sec overlapping portion of the signal. The default MUAP width was set to 25 ms, as used in conventional methods. Finally re-averaging is performed over all MUAPs, using median averaging to reduce the noise caused by interference from other MUAPs. In the next step, temporal energies of all individual MUAPs are investigated. Since, the MUAP energy is a single, physically meaningful parameter that more directly reflects the force-generating capacity of the motor unit, the MUAP exhibiting highest energy content will have dominant effect in corresponding EMG signal. This is why we defined this MUAP as ‘Dominant MUAP’.

In later stage, feature extractions from the dominant MUAPs are carried out. Different time and frequency domain characteristics are adopted to get feature out of MUAPs. A total of 5 morphological features (amplitude, duration, rise time, area, and no. of phases) are extracted from dominant MUAPs as proposed feature of [35]. In order to compare of the performance of the proposed method in terms of accuracy,
the said features are used for classification. A good quality feature is sufficient to classify patterns/signal without the dependence of supervised trained classifiers. Keeping that in mind, we have considered the proposed feature of the said methods only. And the performance is evaluated with simple Knn classifier.

Our proposed DCT based method on dominant MUAP exhibits notable classification accuracy. The classification accuracy is well compared to the proposed RAW EMG signal approach. Our proposed DWT based statistical features outperform the other feature-based methods by significant amount of accuracy. For the feature vector, the maximum of both approximate and details coefficients are considered. And the standard deviations of both coefficients are taken as well. In this work, the standard deviation of the wavelets coefficients is used to identify frequency anomalies in dominant MUAP.

Fig. 3.21: Proposed DWT feature for normal persons and ALS patients (Two class small dataset)

In Fig. 3.21 and Fig. 3.22, the variation of the proposed DWT features for 2 class small database and 2 class large database are shown respectively. And in Fig. 3.23 and Fig. 3.24, the variation of the proposed DWT features for 3
Fig. 3.22: Proposed DWT feature for normal persons and ALS patients (Two class large dataset)

class small database and 3 class large database are depicted respectively. It is found that the proposed feature exhibits higher overall classification accuracy. DWT decomposition of dominant MUAP upto level-6 is also investigated considering all four datasets. The classification accuracy against levels of decomposition is plotted in Fig. 3.25 and Fig. 3.26. The upper plot in Fig. 3.25 is for two class small dataset and lower plot is for two class large dataset. Similarly, the upper plot in Fig. 3.26 is for three class small dataset and lower plot is for three class large dataset. It is found that the accuracy decreases with the increase of decomposition level. Multiple level decomposition breaks the frequency bin to lower region and eventually affecting the time resolution. Therefore, morphological feature value exhibit degraded performance. On the other hand, the length of dominant MUAP is decimated to such level where the chances of getting significant information is uncertain. Therefore, the proposed method utilized level-1 decomposition in order to achieve highest level of classification accuracy.

It is to be noted that the feature dimension is kept extremely low in the proposed
methods, which help in reducing the computational complexity. In case of direct raw EMG based method, proposed DWT Max and DWT ten Max methods require 21.78 sec and 22.62 sec, respectively to perform testing of a dataset considering 25 frames. On the other hand, the proposed MUAP based method require 0.78 sec excluding the MUAP extraction time. For MUAP extraction on an average 15 sec is required. Thus, in comparison to the proposed direct EMG analysis scheme, the proposed MUAP based scheme not only provides better classification accuracy but also offers less execution time.

Tables 3.1 and 3.2 provide comparison of the performance of the proposed method for two class small and two class large database in terms of average accuracy to other methods. Our proposed method gives 100% accuracy for two class small database both in raw EMG and MUAP analysis. It is observed that the accuracy of the raw EMG signal based methods decreases for large database (two class large database), whereas the dominant MUAP based methods still provide high accuracy of 99.5%.

Tables 3.3 and Table 3.4 summarizes the performance of the proposed method
Fig. 3.24: Proposed DWT feature for normal, myopathic and ALS patients (Three class large dataset)

for three class small and three class large database along with the performance of conventional methods. The dominant MUAP based method provides high accuracy for classification of neuromuscular disease from normal control group. The proposed dominant MUAP based approach gives 100% and 98.8% of total accuracy for three class small and large database respectively. Graphical representations of comparison of different proposed methods are summarized in Fig. 3.27, Fig. 3.28, Fig. 3.29, and Fig. 3.30.

3.7 Conclusion

In this chapter, a new method for classification of neuromuscular disease is proposed incorporating EMG signal decomposition, dominant MUAP detection, feature extraction from dominant MUAPs and classifying according to their pathology using kNN classifier. For decomposition, a publicly available decomposition tool “EMGLAB” is used. Since, clinical decomposition methods need to be robust enough to deal with MUAP instability in pathology, an efficient EMG decomposition
Fig. 3.25: Effect of variation of decomposition levels on total accuracy for both two class small dataset (top) and large dataset (bottom)

Fig. 3.26: Effect of variation of decomposition levels on total accuracy for both three class small dataset (top) and large dataset (bottom)
Table 3.1: Performance of proposed classification method experimented on two class small dataset

<table>
<thead>
<tr>
<th>Feature set</th>
<th>Signal</th>
<th>Value of $K$</th>
<th>$Sp$</th>
<th>$SeA$</th>
<th>TAcc in percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>S-AR</td>
<td>Raw</td>
<td>10</td>
<td>83.4667</td>
<td>79.6000</td>
<td>81.5333</td>
</tr>
<tr>
<td>S-DWT</td>
<td>Raw</td>
<td>8</td>
<td>77.4000</td>
<td>77.4000</td>
<td>74.4000</td>
</tr>
<tr>
<td>S-WPT</td>
<td>Raw</td>
<td>7</td>
<td>66.8000</td>
<td>63.2667</td>
<td>65.0333</td>
</tr>
<tr>
<td>S-AR+WPT</td>
<td>Raw</td>
<td>8</td>
<td>82.9333</td>
<td>79.8667</td>
<td>81.4000</td>
</tr>
<tr>
<td>S-AR+DWT</td>
<td>Raw</td>
<td>10</td>
<td>82.6667</td>
<td>80.0000</td>
<td>81.3333</td>
</tr>
<tr>
<td>S-DWT+WPT</td>
<td>Raw</td>
<td>7</td>
<td>77.0667</td>
<td>77.9333</td>
<td>77.5000</td>
</tr>
<tr>
<td>Morphological</td>
<td>MUAP</td>
<td>3</td>
<td>100.00</td>
<td>53.3333</td>
<td>76.6667</td>
</tr>
<tr>
<td>features [35]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time domain:ACR</td>
<td>Raw</td>
<td>3</td>
<td>100.0000</td>
<td>93.3333</td>
<td>96.6667</td>
</tr>
<tr>
<td>Time domain:ZCR</td>
<td>Raw</td>
<td>5</td>
<td>66.6667</td>
<td>60.0000</td>
<td>63.3333</td>
</tr>
<tr>
<td>Freq domain:DCT</td>
<td>MUAP</td>
<td>5</td>
<td>100.00</td>
<td>86.6667</td>
<td>93.3333</td>
</tr>
<tr>
<td>Freq domain:FFT</td>
<td>Raw</td>
<td>1</td>
<td>86.6667</td>
<td>86.6667</td>
<td>86.6667</td>
</tr>
<tr>
<td>Freq domain:FFT</td>
<td>MUAP</td>
<td>3</td>
<td>100.00</td>
<td>86.6667</td>
<td>93.3333</td>
</tr>
<tr>
<td>Freq domain:MDF</td>
<td>Raw</td>
<td>10</td>
<td>86.6667</td>
<td>40.0000</td>
<td>63.3333</td>
</tr>
<tr>
<td>Freq domain:MNF</td>
<td>MUAP</td>
<td>1</td>
<td>66.6667</td>
<td>66.6667</td>
<td>66.6667</td>
</tr>
<tr>
<td>DWT Five Coeff</td>
<td>Raw</td>
<td>1</td>
<td>100.00</td>
<td>100.0000</td>
<td>100.0000</td>
</tr>
<tr>
<td>DWT Fifteen Coeff</td>
<td>Raw</td>
<td>7</td>
<td>100.0000</td>
<td>93.3333</td>
<td>96.6667</td>
</tr>
<tr>
<td>Proposed Method-1: DWT Max</td>
<td>Raw</td>
<td>1</td>
<td>100.0000</td>
<td>100.0000</td>
<td>100.0000</td>
</tr>
<tr>
<td>Proposed Method-2: DWT Ten Coeff</td>
<td>Raw</td>
<td>1</td>
<td>100.0000</td>
<td>100.0000</td>
<td>100.0000</td>
</tr>
<tr>
<td>Proposed Method: DWT Feature</td>
<td>MUAP</td>
<td>1</td>
<td>100.0000</td>
<td>100.0000</td>
<td>100.0000</td>
</tr>
</tbody>
</table>

tool in which automatic decomposition algorithm is used. For identifying dominant MUAPs, highest temporal energy is considered. In order to feature extraction, different time, frequency, morphological and wavelet domain features are extracted. Finally a k-Nearest Neighbor based classifier is applied to classify the EMG signal of neuromuscular diseases group from normal group. It is found that the proposed methods require significantly less computational time to perform the classification task with a very competitive overall classification accuracy. It is therefore highly suitable for neurophysiologists to assist them with the clinical decision.
Table 3.2: Performance of proposed classification method experimented on two class large dataset

<table>
<thead>
<tr>
<th>Feature set</th>
<th>Signal</th>
<th>Value of $K$</th>
<th>$Sp$</th>
<th>$SeA$</th>
<th>TAcc in percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>S-AR</td>
<td>Raw</td>
<td>7</td>
<td>91.7800</td>
<td>48.9400</td>
<td>81.0700</td>
</tr>
<tr>
<td>S-DWT</td>
<td>Raw</td>
<td>10</td>
<td>92.0000</td>
<td>44.8400</td>
<td>80.2100</td>
</tr>
<tr>
<td>S-WPT</td>
<td>Raw</td>
<td>7</td>
<td>90.4467</td>
<td>32.9800</td>
<td>76.0800</td>
</tr>
<tr>
<td>S-AR+WPT</td>
<td>Raw</td>
<td>7</td>
<td>91.7867</td>
<td>48.9600</td>
<td>81.0800</td>
</tr>
<tr>
<td>S-AR+DWT</td>
<td>Raw</td>
<td>10</td>
<td>93.5800</td>
<td>44.7800</td>
<td>81.3800</td>
</tr>
<tr>
<td>S-DWT+WPT</td>
<td>Raw</td>
<td>7</td>
<td>77.0667</td>
<td>77.9333</td>
<td>77.5000</td>
</tr>
<tr>
<td>Morphological features [35]</td>
<td>MUAP</td>
<td>8</td>
<td>94.0000</td>
<td>44.0000</td>
<td>81.5000</td>
</tr>
<tr>
<td>Time domain:ACR</td>
<td>Raw</td>
<td>7</td>
<td>91.1533</td>
<td>45.3000</td>
<td>79.6900</td>
</tr>
<tr>
<td>Time domain:ZCR</td>
<td>Raw</td>
<td>5</td>
<td>92.0000</td>
<td>64.0000</td>
<td>85.0000</td>
</tr>
<tr>
<td>Freq domain:DCT</td>
<td>MUAP</td>
<td>4</td>
<td>94.6667</td>
<td>82.0000</td>
<td>91.5000</td>
</tr>
<tr>
<td>Freq domain:FFT</td>
<td>Raw</td>
<td>7</td>
<td>94.0000</td>
<td>34.0000</td>
<td>79.0000</td>
</tr>
<tr>
<td>Freq domain:FFT</td>
<td>MUAP</td>
<td>3</td>
<td>96.6667</td>
<td>74.0000</td>
<td>91.0000</td>
</tr>
<tr>
<td>Freq domain:MDF</td>
<td>Raw</td>
<td>1</td>
<td>83.3333</td>
<td>30.0000</td>
<td>70.0000</td>
</tr>
<tr>
<td>Freq domain:MNF</td>
<td>MUAP</td>
<td>1</td>
<td>77.3333</td>
<td>36.0000</td>
<td>67.0000</td>
</tr>
<tr>
<td>DWT Five Coeff</td>
<td>Raw</td>
<td>4</td>
<td>92.6667</td>
<td>72.0000</td>
<td>87.5000</td>
</tr>
<tr>
<td>DWT Fifteen Coeff</td>
<td>Raw</td>
<td>1</td>
<td>92.6667</td>
<td>72.0000</td>
<td>85.5000</td>
</tr>
<tr>
<td>Proposed Method-1:DWT Max</td>
<td>Raw</td>
<td>11</td>
<td>96.6667</td>
<td>74.0000</td>
<td>91.0000</td>
</tr>
<tr>
<td>Proposed Method-2:DWT Ten Coeff</td>
<td>Raw</td>
<td>8</td>
<td>97.3333</td>
<td>74.0000</td>
<td>91.5000</td>
</tr>
<tr>
<td>Proposed Method:DWT Feature</td>
<td>MUAP</td>
<td>1</td>
<td>100.0000</td>
<td>98.0000</td>
<td>99.5000</td>
</tr>
</tbody>
</table>
Table 3.3: Performance of proposed classification method experimented on three class small dataset

<table>
<thead>
<tr>
<th>Feature set</th>
<th>Signal</th>
<th>Value of K</th>
<th>Sp</th>
<th>SeA</th>
<th>SeM</th>
<th>TAcc in percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>S-AR</td>
<td>Raw</td>
<td>10</td>
<td>68.1333</td>
<td>76.0667</td>
<td>81.5333</td>
<td>75.2444</td>
</tr>
<tr>
<td>S-DWT</td>
<td>Raw</td>
<td>7</td>
<td>56.4000</td>
<td>75.2000</td>
<td>76.7333</td>
<td>69.4444</td>
</tr>
<tr>
<td>S-WPT</td>
<td>Raw</td>
<td>1</td>
<td>45.6000</td>
<td>56.0667</td>
<td>60.3333</td>
<td>54.0000</td>
</tr>
<tr>
<td>S-AR+WPT</td>
<td>Raw</td>
<td>7</td>
<td>68.5333</td>
<td>74.0667</td>
<td>80.6000</td>
<td>74.4000</td>
</tr>
<tr>
<td>S-DWT+WPT</td>
<td>Raw</td>
<td>7</td>
<td>66.2000</td>
<td>74.8667</td>
<td>81.0667</td>
<td>74.0444</td>
</tr>
<tr>
<td>Morphological features [35]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>MUAP</td>
<td>4</td>
<td>80.0000</td>
<td>53.3333</td>
<td>80.0000</td>
<td>71.1111</td>
</tr>
<tr>
<td>Time domain:ACR</td>
<td>Raw</td>
<td>3</td>
<td>73.3333</td>
<td>93.3333</td>
<td>53.3333</td>
<td>73.3333</td>
</tr>
<tr>
<td>Time domain:ZCR</td>
<td>Raw</td>
<td>7</td>
<td>46.6667</td>
<td>60.0000</td>
<td>53.3333</td>
<td>53.3333</td>
</tr>
<tr>
<td>Freq domain:DCT</td>
<td>MUAP</td>
<td>8</td>
<td>73.3333</td>
<td>86.6667</td>
<td>73.3333</td>
<td>77.7778</td>
</tr>
<tr>
<td>Freq domain:FFT</td>
<td>Raw</td>
<td>1</td>
<td>80.0000</td>
<td>62.0000</td>
<td>20.0000</td>
<td>71.6279</td>
</tr>
<tr>
<td>Freq domain:FFT</td>
<td>MUAP</td>
<td>1</td>
<td>89.3333</td>
<td>80.0000</td>
<td>26.6667</td>
<td>82.7907</td>
</tr>
<tr>
<td>Freq domain:MDF</td>
<td>Raw</td>
<td>1</td>
<td>84.0000</td>
<td>44.0000</td>
<td>26.6667</td>
<td>70.6977</td>
</tr>
<tr>
<td>Freq domain:MNF</td>
<td>MUAP</td>
<td>10</td>
<td>33.3333</td>
<td>60.0000</td>
<td>40.0000</td>
<td>44.4444</td>
</tr>
<tr>
<td>DWT Five Coeff</td>
<td>Raw</td>
<td>4</td>
<td>80.0000</td>
<td>100.0000</td>
<td>80.0000</td>
<td>86.6667</td>
</tr>
<tr>
<td>DWT Fifteen Coeff</td>
<td>Raw</td>
<td>9</td>
<td>73.3333</td>
<td>93.3333</td>
<td>86.6667</td>
<td>84.4444</td>
</tr>
<tr>
<td>Proposed Method-1:DWT Max</td>
<td>Raw</td>
<td>7</td>
<td>73.3333</td>
<td>100.0000</td>
<td>93.3333</td>
<td>88.8889</td>
</tr>
<tr>
<td>Proposed Method-2:DWT Ten Coeff</td>
<td>Raw</td>
<td>3</td>
<td>86.6667</td>
<td>100.0000</td>
<td>86.6667</td>
<td>91.1111</td>
</tr>
<tr>
<td>Proposed Method:DWT Feature</td>
<td>MUAP</td>
<td>1</td>
<td>100.0000</td>
<td>100.0000</td>
<td>100.0000</td>
<td>100.0000</td>
</tr>
</tbody>
</table>
Table 3.4: Performance of proposed classification method experimented on three class large dataset

<table>
<thead>
<tr>
<th>Feature set</th>
<th>Signal</th>
<th>Value of K</th>
<th>Sp</th>
<th>SeA</th>
<th>SeM</th>
<th>TAcc in percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>S-AR</td>
<td>Raw</td>
<td>7</td>
<td>82.7467</td>
<td>48.1600</td>
<td>57.8000</td>
<td>70.8400</td>
</tr>
<tr>
<td>S-DWT</td>
<td>Raw</td>
<td>7</td>
<td>80.8800</td>
<td>46.7600</td>
<td>52.0800</td>
<td>68.2960</td>
</tr>
<tr>
<td>S-WPT</td>
<td>Raw</td>
<td>7</td>
<td>77.3200</td>
<td>33.3400</td>
<td>32.2800</td>
<td>59.5160</td>
</tr>
<tr>
<td>S-AR+WPT</td>
<td>Raw</td>
<td>1</td>
<td>82.7533</td>
<td>48.2200</td>
<td>57.8000</td>
<td>70.8640</td>
</tr>
<tr>
<td>S-AR+DWT</td>
<td>Raw</td>
<td>1</td>
<td>82.1800</td>
<td>46.9800</td>
<td>59.2000</td>
<td>70.5440</td>
</tr>
<tr>
<td>S-DWT+WPT</td>
<td>Raw</td>
<td>1</td>
<td>80.8933</td>
<td>46.9000</td>
<td>51.9600</td>
<td>68.3080</td>
</tr>
<tr>
<td>Morphological features [35]</td>
<td>MUAP</td>
<td>5</td>
<td>86.0000</td>
<td>54.0000</td>
<td>46.0000</td>
<td>71.6000</td>
</tr>
<tr>
<td>Time domain:ACR</td>
<td>Raw</td>
<td>7</td>
<td>84.6667</td>
<td>64.0000</td>
<td>52.0000</td>
<td>74.0000</td>
</tr>
<tr>
<td>Time domain:ZCR</td>
<td>Raw</td>
<td>5</td>
<td>82.6667</td>
<td>30.0000</td>
<td>22.0000</td>
<td>60.0000</td>
</tr>
<tr>
<td>Freq domain:DCT</td>
<td>MUAP</td>
<td>8</td>
<td>90.6667</td>
<td>76.0000</td>
<td>72.0000</td>
<td>84.8000</td>
</tr>
<tr>
<td>Freq domain:FFT</td>
<td>Raw</td>
<td>7</td>
<td>84.0000</td>
<td>28.0000</td>
<td>38.0000</td>
<td>63.0000</td>
</tr>
<tr>
<td>Freq domain:FFT</td>
<td>MUAP</td>
<td>6</td>
<td>88.0000</td>
<td>68.0000</td>
<td>62.0000</td>
<td>78.8000</td>
</tr>
<tr>
<td>Freq domain:MDF</td>
<td>Raw</td>
<td>1</td>
<td>76.6667</td>
<td>28.0000</td>
<td>48.0000</td>
<td>61.2000</td>
</tr>
<tr>
<td>Freq domain:MNF</td>
<td>MUAP</td>
<td>1</td>
<td>62.6667</td>
<td>36.0000</td>
<td>24.0000</td>
<td>49.6000</td>
</tr>
<tr>
<td>DWT Five Coeff</td>
<td>Raw</td>
<td>6</td>
<td>89.3333</td>
<td>70.0000</td>
<td>90.0000</td>
<td>85.6000</td>
</tr>
<tr>
<td>DWT Fifteen Coeff</td>
<td>Raw</td>
<td>3</td>
<td>89.3333</td>
<td>72.0000</td>
<td>84.0000</td>
<td>84.8000</td>
</tr>
<tr>
<td>Proposed Method-1:DWT Max</td>
<td>Raw</td>
<td>17</td>
<td>92.0000</td>
<td>72.0000</td>
<td>82.0000</td>
<td>86.0000</td>
</tr>
<tr>
<td>Proposed Method-2:DWT Ten Coeff</td>
<td>Raw</td>
<td>8</td>
<td>91.3333</td>
<td>74.0000</td>
<td>90.0000</td>
<td>87.6000</td>
</tr>
<tr>
<td>Proposed Method:DWT Feature</td>
<td>MUAP</td>
<td>1</td>
<td>100.000</td>
<td>98.0000</td>
<td>96.0000</td>
<td>98.8000</td>
</tr>
</tbody>
</table>
Fig. 3.27: Classification performance of two class small dataset

Fig. 3.28: Classification performance of two class large dataset
Fig. 3.29: Classification performance of three class small dataset

Fig. 3.30: Classification performance of three class large dataset
Chapter 4

Conclusion

4.1 Concluding Remarks

The main objective of this thesis work is to develop a reliable and computationally efficient classification scheme for neuromuscular disease classification using EMG signal. In order to achieve this target, two different feature extraction schemes are proposed: one directly extract feature from the collected EMG data and the other one perform the task of feature extraction on the MUAPs obtained via EMG decomposition. In both cases, a special care is taken to obtain satisfactory neuromuscular disease classification performance with less computational burden.

In the direct approach, frame by frame analysis is carried out. Most of the direct approaches reported so far treat each frame as a representative of the corresponding class and perform classification for each frame. In the proposed method, instead of relying only on such local information, a global statistics is obtained by averaging the information obtained from a number of consecutive frames. Since only the global information is used for classification, the computational burden of the classifier is drastically reduced. Moreover, it is found that the global statistics, in comparison to local statistics, can provide consistent underlying information from EMG signal. Feature extraction techniques in time, frequency, and the time-frequency domains are then investigated in order to find the appropriate set of features. It is shown that the DWT based proposed features extracted from EMG data can exhibit high within class compactness and between class separability. In the proposed method, both maxima value and ten maxima values of DWT coefficients are utilized, which successively reduce the feature dimension.

In the MUAP based method, the MUAPs are acquired by using a publicly avail-
able EMG decomposition tool (EMGLAB), which reduces the dependency on precise MUAP extraction via needle electrode from motor neurons of muscles. After the decomposition of EMG signal, there may always exist some non-stationary MUAPs which cannot properly represent corresponding class information. However, MUAP based methods available in literature consider extracting features from all available MUAPs, which increases computational burden and decreases the quality of extracted features. Hence, the MUAP that carries most of essential information is to be identified and selected for further analysis. As there is supposedly an attractive connection between motor unit size, force-generation and the energy content of an MUAP, the temporal energy of all MUAPs are computed. The MUAP containing the highest energy content is identified as dominant MUAP and selected for further feature extraction. Some selected conventional time, frequency domain and morphological features are extracted from all of the dominant MUAPs representing its parent EMG signal. Apart from these features, a DWT based statistical feature is proposed which outperforms the quality of existing features.

Classification performance is tested both in small and large datasets considering two and three class problems. The proposed methods exhibit very high classification accuracy in comparison to some of the existing methods.

4.2 Major Contribution of This Thesis

- Instead of local information computed from a frame of EMG data, the global statistics is considered, which is computed by averaging the information obtained from some consecutive frames. This not only reduces the computational complexity but also helps in obtaining more consistent feature characteristics.

- Instead of all DWT coefficients, in the proposed method, some statistics of the coefficients are considered, which drastically reduces the computational burden.

- The proposed methods can efficiently handle both the two class and three class problems.

- The new idea of dominant MUAP selection offers advantage of ignoring MUAPs those are not capable of properly representing the corresponding class.
It also in turn reduces the computational burden.

- The extracted wavelet based features offer a very high within class compactness and between class separability.

- Because of the robustness of the proposed feature, use of a simple KNN classifier can provide excellent classification accuracy.

- The proposed features provide a very high classification accuracy in comparison to some of the existing features such as different time and frequency domain features, autoregressive power spectrum, wavelet packet energy and DWT statistical features.

- Both small and a very large datasets are used for performance evaluation.

### 4.3 Scope and Future Work

- In the proposed classification method, like most of the EMG analysis schemes, needle EMG signal is used. One possible future work could be to test the proposed dominant MUAP based method on surface EMG signal extracted from muscles. Since surface EMG signal is acquired from surface of the muscles, it would be difficult to obtain all MUAPs precisely. As a result, it is expected that instead of working with all MUAPs, the proposed dominant MUAP based method would provide better performance, thereby reducing the dependency on invasive needle EMG extraction process.

- Another future work could be to incorporate speech like characteristics of the EMG recordings. During EMG recording, a characteristic crispy repetitive sound is generally heard when the tip of the needle is close to some muscle fibers. Hence, it is expected that by applying some speech signal processing algorithms on EMG data, distinguishable features may be obtained.

- There are some other neuromuscular diseases, such as parkinson disease and muscle fatigue, which are getting much attention by a lot of researchers nowadays. It would be very interesting to investigate the performance of the proposed classification scheme on databases collected from these types of patient groups.
Appendix A

EMGLAB: EMG Decomposition Method

EMGLAB provides a convenient graphical interface as well as a number of automatic procedures for decomposing and inspecting EMG signals. The computer screen is divided into four panels that show a segment of the EMG signal, the templates of the identified MU spikes, the discharge patterns of the identified MUP trains, and a close-up of the signal for resolving superimpositions. Manual decomposition functions can be performed using the graphical interface. For example, new templates can be formed by dragging spikes from the signal panel to the template panel, and spikes in the signal panel can be identified either by dragging specific templates over them or by shift-clicking on them to have the program determine the best-fitting template. Graphical commands are also available for undoing identifications; deleting, reordering, and merging templates; deleting points from the discharge panel; and selecting and adjusting the template configuration in the close-up panel.

The decomposition process typically proceeds in the following way. First, the program reads in the initial 2 sec of the EMG signal and automatically creates templates for all the spikes that occur at least three times with a high-degree of similarity. Then, the program tries to automatically classify the remaining spikes in this 2 sec interval using template matching. Depending on the complexity of the signal, these automatic procedures may or may not achieve a full decomposition. The signal is then inspected manually to complete the decomposition and verify the results. This may require several passes back and forth through the data. Once this initial 2 sec interval has been fully decomposed, the next 2 sec are read in and analyzed using the existing set of templates, and so on. The accuracy of the decomposition
is assessed subjectively in two main ways. The first way is by inspecting the lower trace in the signal panel (Fig. A.1). This trace shows the residual that remains after the templates of the identified MUs have been subtracted from the signal. A small residual indicates a good fit between the templates and the signal, whereas a large residual indicates an incomplete or incorrect identification. The second way of assessing accuracy is by inspecting the identified discharge patterns (Fig. A.3). Full, regular patterns provide confidence that the decomposition is correct, whereas gaps, extra discharges or uneven intervals are signs of possible decomposition errors. The goal is to achieve a flat residual signal and smooth, regular discharge patterns, which together indicate a complete and accurate decomposition.

Fig. A.1: Signal panel, showing a segment of the EMG signal (top trace) and the residual signal after template matching (bottom trace)

**Signal Panel**

The signal panel makes it possible to scroll through and inspect the EMG signal in close detail (Fig. A.1). The signal is displayed in its entirety, rather than just the spikes that exceed a certain threshold. This makes it easier to see the temporal relationships between spikes and spike components and decreases the chances of overlooking small but important signal details.

The signal can be displayed either unfiltered or after digital high-pass filtering. For decomposition, a 1kHz high-pass filter is usually used. This flattens the signal baseline and sharpens the MUPs into narrow spikes, making them easier to detect.
and distinguish. For averaging, the unfiltered signal is used in order to capture the entire MUP waveforms.

**Templates**

The default template length is 12.8 ms, which is long enough to hold most MUP spikes, including small outlying components. For MUPs with more than one distinct component, separate templates can be created for each component to reduce residual noise from inter-component jitter. Slow changes in spike shape are tracked by periodically reaveraging the templates. Re-averaging is performed over a 2 sec window, using median averaging to reduce the noise caused by interference from other MUPs. An example of MUAP templates are shown in Fig. A.2

![Signal: N2001C01BB70.hea; Annotation: untitled; Chan 1; unfiltered](image)

**Fig. A.2:** Template panel, showing the templates of the identified MUAPs

When a template is dragged to a spike in the signal panel, the program automatically aligns the template with the spike to minimize the resulting residual. From the operator’s point of view, the template seems to snap into place. The alignment is performed to sub-sampling-interval precision (0.01 ms) using interpolation to eliminate residual error due to time quantization.

**Discharge Patterns**

In many signals, timing information is helpful for decomposition. During steady and slowly changing contractions, MUs usually discharge with fairly regular inter-
discharge intervals (IDIs) (Fig. A.3). This regularity may not be apparent at the beginning of a decomposition when only a few scattered discharges have been identified, but it becomes clear as more of the discharge pattern is filled in. Once several of a MU’s adjacent discharges have been identified it becomes possible to estimate the mean IDI, which, in turn, helps determine where to look for the subsequent discharges. As the discharge pattern becomes completely filled in, discrepancies in the pattern stand out clearly as indicators of possible errors.

![Discharge panel](image.png)

**Fig. A.3:** Discharge panel, showing the identified discharge times of each MU. The vertical lines indicate the interval displayed in the signal panel.

To help locate the discharges of a selected MU, the program displays bars in the signal panel that indicate the MU’s expected discharge times (Fig. A.1). Bars are shown before and after each of the MU’s already identified discharges. Their locations and lengths are based on the MU’s estimated firing statistics over a 2 sec window. Since the identified discharge pattern might contain misses or false positives, IDIs that are much longer or shorter than the median IDI are excluded from the statistics calculation.

It should be mentioned that not all signals exhibit the same degree of discharge regularity. MUs near their recruitment thresholds can fire intermittently, and some individuals, and perhaps some muscles, exhibit less regularity and less fluctuation synchrony than others.
Resolving Superimpositions

One of the most difficult aspects of EMG decomposition is resolving superimpositions. In a signal with 15 active MUs, each firing at a mean rate of $15Hz$, a third of all discharges will occur within $1ms$ of another discharge, and there will be about 15 instances per second in which four different discharges occur within the same $4ms$ interval. When this happens, the individual MUPs sum together to produce a superimposition. Depending on the precise timing, superimpositions can range in complexity from partial ones in which the individual constituents are still largely recognizable, to full ones in which the constituents are unrecognizable because of constructive or destructive interference.

EMGLAB is able to resolve many superimpositions during the automatic template-matching phase by identifying larger spikes and "peeling them off" to reveal smaller spikes in the residual. Superimpositions that are not resolved in this way must be resolved manually. This is done in the closeup panel (Fig. A.4), which displays the superimposition at an expanded scale and allows different sets of templates to be selected and adjusted to find the best fit. The program can also be instructed to determine the optimal alignment for the selected set of templates. This turns out to be a difficult problem because of the large number of possible ways that the templates can be aligned. The algorithm used by EMGLAB essentially considers every possible alignment, using a branch-and-bound approach to try the most likely alignments first and to stop once it can be determined that none of the remaining alignments could do better than the ones already tried. Interpolation is used to reduce time-quantization errors. The algorithm is able to resolve complicated superimpositions involving six or more MUPs quite efficiently, even in cases of destructive interference and different sized MUPs.

The templates involved in a superimposition can usually be determined, or at least narrowed down, by inspecting the discharge patterns to see which MUs are expected to fire at the time in question. The program does not attempt to try all the different possible combinations of templates, because that would be prohibitively time consuming. It would also often lead to the wrong result. The reason is that it is often possible to find incorrect sets of templates that provide a better fit to the superimposition than the correct set does. The incorrect sets usually include too
Fig. A.4: Close-up panel, showing a segment of the EMG signal at an expanded scale (top trace), a configuration of templates selected to match it (bottom traces and numbers), the sum of the templates (top dotted trace), and the residual (center trace). All the time scales are in seconds from the start of the signal.

many templates, which, because of the extra degrees of freedom, can be aligned to match some of the background noise as well as the superimposition itself.

**MUAP waveforms**

The MUAP waveform recorded by a monopolar electrode is a temporal record of the electrical events that take place during a MU action potential. In particular, the onset, spike, and terminal wave of the MUAP mark, respectively, the initiation of the action potential at the endplate, its propagation past the electrode, and its termination at the muscle/tendon junction. Thus, by measuring the relative latencies between these MUAP features it is possible to estimate the architectural organization of the MU. This approach provides one of the few available methods for studying the architectural organization of individual MUs in vivo.

To study muscle architecture, it is important to obtain an accurate estimate of the complete MUAP waveform, including its onset and terminal wave. These features tend to be low-frequency, and so are much better seen in the unfiltered signal than in the high-pass filtered signal used for decomposition. They are also often quite small in amplitude, and so signal averaging is usually needed to detect them.
reliably. EMGLAB uses an averaging algorithm that estimates and subtracts out the effect of the interference from the other MUAPs to achieve signal-to-noise ratios much higher than those obtained using simple averaging. In this way, EMGLAB is able to obtain MUAP averages acceptable for architectural analysis from 10 or 20 s long epochs, even in fairly complex and noisy signals. It should be noted that the MUAP waveform is not perfectly constant from discharge to discharge. Rather, it varies somewhat in shape and duration, due primarily to fluctuations in muscle-fiber conduction velocity.
Bibliography


