Detection of Amyotrophic Lateral Sclerosis using Signal Processing and Machine Learning

by

Zohair Mehtab Ali
19341013

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science

Department of Computer Science and Engineering
Brac University
April 2019

Declaration

It is hereby declared that

1. The thesis submitted is my/our own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Zohair Mehtab Ali
19341013

Approval

The thesis/project titled "Detection of Amyotrophic Lateral Sclerosis using Signal Processing and Machine Learning" submitted by

1. Zohair Mehtab Ali (19341013)

Of Summer, 2019 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on April 25, 2019.

Examining Committee:

Supervisor:
(Member)

Dr. Mohammad Zavid Parvez
Assistant Professor
Department of Computer Science and Engineering
Brac University

Program Coordinator:
(Member)

Dr. Jia Uddin
Associate Professor
Department of Computer Science and Engineering
Brac University

Head of Department:
(Chair)

Dr. Md. Abdul Mottalib
Chairperson
Department of Computer Science and Engineering
Brac University

Abstract

Electromyography(EMG) signals provide significant information for the diagnosis of neuromuscular disorders like Amyotrophic Lateral Sclerosis(ALS) which is a form of Motor Neuron Disease(MND). Due to the stochastic nature of EMG signals different preprocessing and feature extraction techniques need to be applied in order to extract useful information from the raw noisy signals. Time-Frequency analysis and EMG Decomposition are two of the widely implemented techniques for feature extraction from EMG signals. However, due to extrinsic and intrinsic artifacts any one feature extraction technique alone does not provide enough information in order to show a consistent performance of classification across a variety of dataset. EMG signal data set acquired from different sources provide varying outcome when passed through the same classification technique. This is a major problem while creating software which is able to perform automated classification and analysis of EMG signals on a wide variety of data set with minimum human intervention. This paper proposes a method for classification of ALS based on evaluation of multiple features extracted from three domains of EMG signal: time domain representation, frequency domain representation and Muscle Unit Action Potential(MUAP) waveform acquired via EMG decomposition of the signal. 43 features were evaluated using feature selection techniques like chi-squared test and recursive feature elimination. Our experimental results show that amplitude, duration and area of the MUAP waveform estimated for each motor unit, inter-spike-intervals of the motor units, variance, zero crossings, zero lag of autocorrelation, waveform length and slope sign change of the time domain representation, average spectral amplitude, total power, variance of central and mean frequency from feature domain representation of the signal provides the best accuracy at an average rate of 85%, a true positive rate(TPR) of 86% and a false positive rate(FPR) of 20% approximately.

Keywords: Signal Processing; Machine Learning; Amyotrophic Lateral Sclerosis; EMG Classification

Dedication

Dedicated to all the scientist and engineer whose untiring and selfless efforts have brought us to where we are today.

Acknowledgement

Search for knowledge has been one of the greatest influences of human advancement since the Ptolemaic Kingdom in ancient Egypt. The library of Alexandria, although lost is a symbol of the infinite capabilities of human mind. Search for knowledge is a search for truth and science is a pillar of truth.

This thesis would not have been possible without the help of a lot of people. I would like to thank all of them for their invaluable support and advice.

First of all, I would like to thank my supervisor Dr. Mohammad Zavid Parvez for all the support and advice he has provided me throughout this research and made this thesis possible. I would also like to thank my parents, sibling, family members and friends for their unconditional love and support. I would like to thank the Creator whose blessings made this research possible.

Most of all, I would like to pay my gratitude to the entire scientific community for leading humankind towards truth and serenity.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

This chapter presents a brief overview of the research by introducing the problem followed by the aims, objectives and the outline for conducted research.

1.1 Overview

Amyotrophic Lateral Sclerosis(ALS), also known as Lou Gehrig's disease or motor neuron disease(MND) is a neuromuscular disorder caused due to the deterioration of motor neurons in the central nervous system which causes loss of voluntary movement[67]. The term Motor Neuron Disease(MND) is generally used to denote ALS which is one form of MND and is sometimes used for the entire group. While MND refers to a collection of diseases there are numerous other kinds of diseases that affects the motor neurons of the nervous system. Therefore, diseases related to motor neurons are called 'Motor Neuron Disorders' which includes MND and other kinds of diseases that affect motor neurons like Spinal Muscle Atrophies [57]. Here, the term MND will refer to ALS and not other motor neuron disorders. The effected regions generally are bulbar, thoracic, abdominal and limb muscles [22]. The exact cause of MND is still unclear but some of the factors that are considered to cause MND are genetic and environmental factors, ageing, exposure to chemicals, fractures and injuries, stress, viral infection, etc. From a neurological viewpoint the causes of MND are Faulty Scaffolding, disruption in transportation of RNA - a protein molecule responsible for transmitting impulses from motor neurons to different parts of the body, formation of abnormal clumps of protein molecule in motor neurons, formation of Oxygen free radical which is a toxic waste near the cells, abnormal functioning of mitochondria and due to glia cells near the motor neurons[27], [50], [68], [102]. About eighty percent of patients of MND are affected by ALS. The symptoms of ALS tends to start from the muscles of hand and feet. The muscles initially seems to become stiff and weak which is accompanied by difficulty in swallowing and speech [70].
Currently there is no cure for MND and the life expectancy of a patient with MND is from two to five years although there might be some exception[40], [50]. The diagnosis of MND is usually done by recording the electrical impulses created within the nervous system by motor neurons. EMG, Magnetic Resonance Imaging(MRI), Transcranial Magnetic Stimulation and Surface EMG(sEMG) are some of the signal acquisition techniques used for recording the activities of motor neurons of a nervous system. Currently the disease has no definite means of identification but is rather

identified by the process of eliminating the possibilities of other diseases. This task is not only tedious but also daunting for the patients and their family members. Therefore, early detection of MND will not only minimize the suffering for patients undergoing several tests just for the detection of this disease but will also provide a platform to invent a method for preventing the disease from spreading.
Electromyography(EMG) is an electrodiagnostic technique which is used for evaluating and recording the electrical activity produced by different skeletal muscles of the body [89]. EMG is a very useful device for detection of different myopathic and neuropathic diseases like ALS as it can directly measure the electrical current of different regions of the body. However, due to a lot of noise information (also known as artifacts) because of external factors like ambient noise, inherent noise in electronics equipment, motion artifact, muscle cross-talk, baseline shift, nearby electrical equipment like cell phones or lights, etc. [103]. As are result the EMG signal needs to be filtered, smoothed and preprocessed before use in order to obtain important information that might be useful for classification of different diseases like ALS. Different kinds of signal processing techniques have been implemented in order to extract useful information from EMG signals which might be relevant for the diagnosis of ALS.

1.2 Problem Statement

Time-Frequency Analysis(TFA), Discrete Wavelet Transform(DWT) and EMG Decomposition Algorithm(EDA) are three of the commonly used techniques for the classification of Electromyography(EMG) signals of neuropathic, healthy and myopathic subjects[18], [26], [38], [56], [61], [77], [90], [91], [107]. However, the performances of these techniques vary with respect to the dataset using which the classification is being performed. As for example, according to Subasi[90], classification of myopathic, neuropathic and healthy subjects provide a performance accuracy of 96.75%, 95.1% when classified with SVM and KNN algorithms respectively. The data was acquired from biceps brachii of the muscle using a concentric needle electrode with a bandpass filer of 5Hz to 10KHz and a sampling rate of 20KHz sampled for a duration of 5 seconds. The force applied by the muscle was 30% Maximum Voluntary Contraction(MVC) under isometric conditions. The data was later passed through DWT using Daubechies4(db4) wavelet and 5 level of decomposition. The features extracted were Mean of the absolute Value(MAV), Standard Deviation(SD) and Average power(AP) of each sub-band coefficient and the Ratio of the absolute mean values(RMAV) of adjacent sub-bands. Total number of features extracted from each signal were 23(3x6=18 for 6 wavelets and 1x5=5 for 5 wavelets). In comparison to that, when the same EMG technique was used on a datset provided by Nikolic M. at Faculty of Health Science, University of Copenhagen in 2001[41] which also uses similar parameters provide a performance accuracy of around 60% and 70% for SVM and KNN classifiers respectively. This clearly shows a variation in performance of different EMG classification algorithms when subjected to different dataset. Furthermore, our study shows that the accuracy of TFA ranges around 70-100% when conducted under some experimental setup whereas its accuracy drops down to a range of 50-70% and 70-90% for real clinical and simulated EMG signal data when the classification is performed using open data. Furthermore, the accuracy of DWT fluctuates around 95-97% for some experimental data whereas

it fluctuates around 60-80% and 70-75% for real clinical and simulated EMG signals respectively when conducted with open data. Similarly, the accuracy of EDA ranges around 60-70% and 70-75% for real and simulated signals respectively. Such variation of performance acts as a barrier to the development of an automated and adaptable software capable of consistent performance on a wide variety of dataset. Moreover, redundant feature selection is one of the main problems in classification of EMG and a careful selection of feature is required based on the problem space in order to perform efficient classification[28], [71], [81], [87]. Each of the three aforementioned analysis techniques have quite a large number of possible candidate features which is likely to carry important information regarding the problem in hand. As for example, Phinyomark et al. [81] evaluated 37 time and frequency domain features for classification of wrist and hand movements and Boostani et al.[44] evaluated 19 features for the control of a prosthetic hand. Moreover, even though some of the classification techniques show a high accuracy[36], [64], [77], [78] using a single domain feature extraction techniques the algorithm is unable to show a consistent performance when the algorithms are implemented on variety of dataset.

1.3 Aim of Study

The main purpose of this study is divided into two of the following parts:

1.3.1 Review of different EMG feature extraction and classification techniques and their limitations during the development of a general purpose application software

This paper performs a systematic comparison of TFA, DWT and EDA for classification of ALS and healthy subjects based on the variation of dataset. The variation of dataset are based on three factors: change of data repository which shows the variation of performance for dataset acquired from two different clinical environments, change of size of input data for training a classifier which shows the adaptation capability of the algorithm to varying size of training data and change of signal type which shows the variation of performance for dataset acquired for two different types of signal like simulated EMG signals and Real EMG signals.

1.3.2 A new Multi-Domain approach for the classification of Amyotrophic Lateral Sclerosis

The main purpose of this paper is to extract 17 time domain features, 8 time-frequency domain features, 15 features obtained from the morphology of MUAP waveform and 3 features obtained from the firing pattern of motor units calculated using EMG decomposition. Two candidate best-feature-sets are created using two feature selection techniques(FST): chi-squared Test(CST) and recursive-feature-elimination(RFE). Finally the two different feature sets are individually classified using the Random Forest Algorithm(RFA) which is an ensemble supervised learning method for classification and K-Nearest Neighbor(KNN) classifier which is a non parametric clustering method used for classification and regression. The hyper

parameters such as the optimum number of decision trees(RFA) or neighbors(KNN) are selected using Particle Swarm Optimization(PSO) which is a is a computational method that optimizes a problem by iterative attempt to improve a candidate solution with regard to a given measure of quality. Performance of the classifiers were then evaluated using Receiver Operating Characteristic(ROC) curve which is a graphical plot that illustrates the diagnostic ability of a binary classifier system as its discrimination threshold is varied.

1.4 Thesis Outline

1. Chapter 2 provides a brief discussion starting from the human nervous system up to motor neurons and finally about ALS, its type and causes in section 2.1. Section 2.2 contains a review of different machine learning algorithms currently used for classification and regression. It also provides a brief review on existing systems used for the diagnosis of ALS in section 2.3. Section 2.4 of this paper provides a brief introduction to EMG followed by a brief review of the existing EMG classification systems currently being implemented in clinical sectors and in research laboratories for diagnosis of ALS.

2. Chapter 3 provides a detailed description of the procedures followed in order to prepare the dataset in section 3.1. Rest of the chapter is divided into two parts. Section 3.2 provides a detailed description about the system architecture developed in order to compare the performance of different EMG classification algorithms as mentioned in section 1.3.1. Section 3.2 provides the detailed description of the system architecture developed in order to perform classification of ALS using multi-domain feature extraction and selection as mentioned in section 1.3.2.

3. Chapter 4 is divided into two main sections. Section 4.1 provides the experimental results obtained from the system architecture mentioned in section 3.2. It also performs a detailed analysis of the results found. Similarly, section 4.2 provides the experimental results obtained from the system architecture mentioned in section 3.3 and performs a detailed analysis on the obtained results.

4. Chapter 5 concludes the paper with general remarks and a brief discussion on the limitations and scope of improvement.

Chapter 2

Literature Review

2.1 The Nervous System and Motor Neuron Disease

2.1.1 The Nervous System

Nervous system is the part of an animal that detects change in environment that leaves a direct or indirect impact on the body(like touch, perfume, etc.) and then works in coordination with Endocrine system of the body in order to respond to such events [58]. It helps to coordinate our actions and process information also known as external stimuli which it receives from the outside world. In vertebrate, the nervous system is mainly divided into two parts: Central Nervous System(CNS) and the Peripheral Nervous System(PNS). CNS is the main processing unit of the nervous system and PNS are bundles of fibers with neurons that help to transmit information from different parts of the body to the Central Nervous System and vice versa.

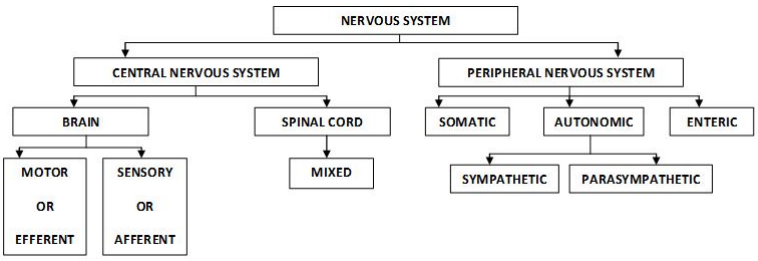

Figure 2.1: Classifications of the Nervous System

Figure 2.1 shows classification of the nervous system in vertebrates.

Central Nervous System(CNS)

It is the main and largest portion of the nervous system consisting of the brain and spinal cord. It integrates information that it receives in form of stimuli and coordinates action of bilaterally symmetric animals. There are different kinds of cells in brain and spinal cord of the Central Nervous System that manages different parts of the body and CNS. Some of the cells are discussed below.

1. **Astrocytes:** It is a kind of glial cell that removes chemicals and protein that might be harmful for the neurons of CNS. Glutamate is an example of Astrocyte which kills neuron by a process called excitotoxicity. They produce proteins known as neurotrophic factor.

2. **Microglia:** They are immune cells in the body that responds to injured part of a body by removing dead or dying cells.

3. **Oligodendrocyte:** It is a cell that wraps around axons of a neuron in order to transmit information in form of electrical signals at a faster rate. Normally electrical messages are transmitted via axons at a speed of one meter per second but with Oligodendrocytes wrapped around the axons, information is transmitted at a speed of one hundred meter per second.

Peripheral Nervous System(PNS)

The Peripheral Nervous System consists of ganglia and nerves which are bundles of axons that connects to the different parts of body at one end and the Central Nervous System in other. It serves as an information relay system between body and CNS [69].The Peripheral Nervous system is volatile to toxicity and mechanical injury as it is not protected by skull and the vertebral column.

Brain

It is the central organ of CNS which helps in processing of information, coordination of activities and decision making of an animal. It mainly consists of cerebrum, cerebellum and the brain stem.

Spinal Cord

Spinal cord is the tube like bundle of tissues in an animal that extends from medulla oblangata of the brain to lumbar region of the vertebral column. It acts as the pathway for transfer of information between the body and brain. The transfer of information takes place by interconnected neurons which is made up of cell body, axons and dendrite.

Somatic Nervous System

The Somatic Nervous System is located in PNS which consists of the sensory nervous system and somatosensory system. It consists of twelve pairs of Cranial Nerve that transmits somatosensory information from body and head [85].

Autonomic nervous System

it is located in the Peripheral Nervous System and largely controls different functions of internal organs of the body like heart beat, digestion,etc [15]. It is also responsible for flight-or-fight response of the body. Most of the functions of Autonomic Nervous System takes place unconsciously and is regulated by the hypothalamus.

Sympathetic Nervous System

It consists of neurotransmitter like Norepinephrine and Epinephrine which increases bodily functions like heart rate and decreases the functions that are non critical for survival like digestion. It is activated during the flight-or-fight response of the body [109]. It controls activities that are responsible for survival. It functions involuntarily i.e. we cannot control activities of the sympathetic nervous system. As for example, the shaking of knee in the state of fear, increase of heart rate, etc.

Parasympathetic Nervous System

Parasympathetic Nervous System helps the body to function in the state of rest or digestion. When an animal is not in a state of flight or fight then the activity of sympathetic nervous system decreases and the parasympathetic nervous system increases.As a result, there is an increase and salivation and activities related to digestion [54]. The parasympathetic nervous system can be controlled voluntarily like urination and defecation.

Enteric Nervous System

Enteric Nervous System is a part of the autonomic nervous system which is located at the digestive tract. It functions on its own without any input from the Sympathetic or Parasympathetic Nervous System but can still interact with rest of the body [88]. It has various functions related to digestive system.

Afferent Neurons

Afferent or Sensory Neurons in a nervous system are the neurons that receive input from the outside world in form of external stimuli like light, sound, etc. and converts it into electric signal by a process called sensory transduction in order for the brain to process the information [25]. The electric signals are also known as action or graded potential. As for example, some neurons respond to tactile stimuli like heat, pressure, etc. and activate motor neurons in order for a muscle to contract or expand as a response.
The cell bodies of Afferent neurons are located in dorsal ganglia of the spinal cord which receives external stimuli from the body and transmits it via Afferent nerve fiber which is a bundle of axons to the brain [76]. Some Afferent neurons receive external stimuli and hence are known as extoreceptors while others receive internal stimuli from the body and hence are known as interoreceptors. There are different kinds of Afferent Neurons based on their functionalities. They are as follows

1. **Olfactory Receptor Neuron:** Some of the Afferent neurons receive external stimuli like molecules of odor from the air which later undergoes transduction and forms the perception which we known as 'Scent' [35].

2. **Gustatory Receptor Neuron:** Some of the Afferent neurons receive external stimuli from molecules of chemical and undergoes transduction in order to form the perception of 'Taste'.

3. **Photo Receptor Neuron:** These Afferent neurons receive external stimuli in form of light or electromagnetic radiation which undergoes transduction in order to form the perception of 'Vision' [74].

4. **Auditory Neuron:** these Afferent neurons receive waves generated from vibrating air molecules and undergoes mechano-electrical transduction in order to form the perception of 'Sound'. These transduction is performed using the hair cells of ear which hyper-polarizes and de-polarizes due to movement and thus releases neurotransmitters.

5. **Thermoreceptor Neuron:** These afferent neurons converts an external stimuli in order to form the perception of 'Warm' or 'Cold'. The mechanism by which transduction occurs is still unclear.

6. **Mechanoreceptor Neuron:** These neurons respond to mechanical forces such as pressure or distortion and converts it into action potential in order for the brain to perceive 'Touch'.

7. **Proprioceptor Neuron:** These neurons provide spatial information about different parts of the body to the brain.

8. **Nociceptor:** These neurons respond to damaging stimuli such as damage to tissues in the body and forms the perception of 'Pain' [100]. Apart from that it also responds to stimuli that might represent a potential damage to a cell. They are found both in internal and external organs of the body. Nociceptors are of three kinds. They are: thermal, mechanical and chemical.

Efferent Neuron:

An efferent or motor neuron receives information from the Central Nervous System and transmits it to an effector organ of the body in order to generate an action like movement. These neurons instruct the muscles to perform an action. These neurons are mainly of two types: Upper and Lower Motor Neurons [55]. Upper motor neurons are located in motor cortex of the brain while the lower motor neurons are located near spinal cord of the body. Information is transmitted from upper to lower motor neuron which later affects the effector organs of the body.

2.1.2 Motor/Efferent Neuron

Motor Neurons of the nervous system helps in voluntary and involuntary movement of organs like hand, facial skin, ventilation system of the body, etc. They are generally located in the motor cortex, brain stem and spinal cord. These neurons are mainly of two types. They are:

1. Upper Motor Neuron located in cerebral cortex of the brain.

2. Lower Motor Neuron located in ventral horn of the spinal cord's gray matter and cranial nerves of the brain stem.

Upper motor neurons transmit information in form of electro-chemical impulses to the lower motor neurons directly or through inter-neurons which is later forwarded from the spinal cord to the effectors.

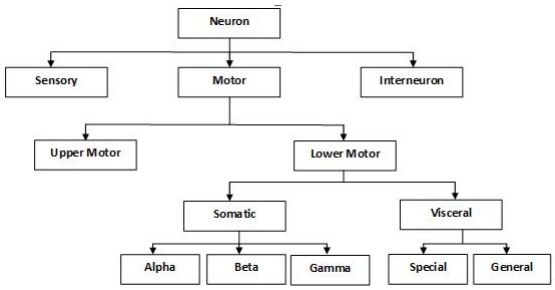

Figure 2.2: Classification of motor neurons

Figure 2.2 displays different types of motor neurons of the nervous system. The term 'motor neuron' generally refers to the lower motor neurons.

Upper Motor Neuron

Betz cells are the largest pyramidal cell that make up the primary motor cortex of the precentral gyrus. Upper motor neurons or the Efferent neurons are located V of the primary motor cortex whose axons travel down to form the corticospinal tract in medulla oblangata on each side of the spinal cord.

Lower Motor Neuron

Lower Motor Neurons function as the interconnection between Upper Motor Neurons and the muscles or effector organs. They are located anterior gray column, anterior nerve root or the nuclei of cranial nerve in the brain stem and is responsible for all voluntary movement of the face like movement of eyes, face, tongue, chewing, vocalization and swallowing.

Mechanism for muscle movement

Glutamate released from Upper Motor neurons trigger depolarization of the Lower Motor neurons located in the anterior gray column. As a result an action potential travels through the axons of the Lower Motor neurons to the neuromuscular junction where Acetylcholine is released. It carries the signal across synaptic cleft to the post synaptic receptors of the muscle cell membrane. As a result muscle contracts which results in movement.

Somatic Motor Neuron

The Somatic Motor Neuron originating in the Central Nervous system is involved in locomotion. They are connected to skeletal muscles via their axons or efferent nerve [47]. They are responsible for voluntary movement of the body which makes up the Somatic Nervous System of PNS.

Alpha Motor Neuron

Alpha Motor Neuron is a type of Somatic Motor Neuron that is responsible for innervating the extrafusal muscle fibers [8]. These fibers are the main force generating component of the muscles. A single motor neuron with all its connected muscle fiber

is known as a 'Motor Unit'. These Motor Units, based on physiological classification may be of three types. They are:

1. **Slow Motor Unit:** These motor units are responsible for controlling small muscles. They are very resistant to fatigue as they contract very slowly and hence provides small amount of energy [5]. These muscles as a result sustain contraction is used to keep the body upright. These muscles gain their energy via oxidative means and hence requires energy. These muscle fibers are also called 'Red fibers'.

2. **Fast Fatiguing Motor Unit:** These motor units are responsible for controlling large muscles. They are very less resistant to fatigue but can apply a large amount of force. These muscles are used for task that require short burst of energy like running, jumping, etc. These muscles gain their energy via glycolytic means and hence do not require oxygen. These fibers are also called 'White fibers'.

3. **Fast Fatigue Resistant Motor Unit:** These motor units stimulate medium sized muscles that is much more resistant to fatigue than Fast Fatiguing Motor Unit but cannot react as fast as them. These muscles provide a greater force than the red fibers and use both oxidative and glycolytic means in order to acquire energy.

The number of Alpha Motor Neuron is directly proportional to the amount of fine motor control in the body. As for example, the number of Alpha motor neurons involved to move a finger is greater than that of quadriceps.

Location of Alpha Motor Neuron

1. **Spinal Cord:** The cell bodies of Alpha motor neurons are located in ventral horn of the spinal cord and have the ability to synapse with around one hundred and fifty muscles fibers on an average [58]. Higher segments of spinal cord contains muscle that stimulates higher parts of the body like segment C5 to C7 contains Alpha motor neurons that stimulates the muscle of biceps. on the other hand, muscles of leg are controlled by Alpha motor neurons located in the lower segments of spinal cord like S1 and S2. In spinal cord, Alpha Motor neurons are located in Lamina IX region of gray matter in the spinal cord.

2. **Brain Stem:** In brain stem the Alpha motor neurons are located within a cluster of cells called the nuclei along with other neurons like neurons of cranial nerve. Not all cranial nerve motor contains Alpha motor neuron. It is also found in medulla, pons and mid-brain.

Renshaw Cells in Alpha Motor Neuron

These cells are inhibitory interneurons found in the gray matter of spinal cord. These cells are kept informed of how vigorously the Alpha motor neurons are firing as they receive excitatory stimuli from the Alpha motor neurons as they emerge. It is also connected to the pool of corresponding Alpha motor neurons as it synapses with one of the neurons from the pool and thus sends inhibitory stimuli. In this way

the Renshaw cells represent a negative feedback mechanism for the Alpha Motor neurons [83].

Inputs to Alpha Motor Neuron

1. **Upper Motor neuron:** Alpha motor neurons receive input from the Upper motor neurons via several path including but not limited to Corticonuclear, Corticospinal and Rubrospinal tracts [110]. The Corticospinal and Corticonuclear tracts are involved in information transfer from upper to the lower motor neurons regarding voluntary movement.

2. **Periphery:** The sensory input to Alpha motor neurons is extensive and has its origin in the Golgi tendon organs and other sensory neurons of the periphery. It creates a complex neural circuit which is responsible for reflex like knee jerk during fear.

3. **Interneurons:** The most extensive input to Alpha motor neuron is from local interneurons of the spinal cord.

Outputs from Alpha Motor Neuron

1. Like other neurons Alpha motor neurons transmit information in form of electrochemical impulses known as action potentials. Action potentials are rapid changes in electrical activity that is transmitted from cell body of the neuron to its axon [3].

2. In order to increase the speed of information transmission, these neurons have large diameters and are myelinated by Oligodendrocytes and Schwann Cells.

3. Oligodendrocytes myelinate the portion of axon of Alpha motor neurons that lies in the CNS while Schwann Cells myelinate the the axons that lie in PNS.

4. The transition between information of CNS and PNS takes place at the level of Pia meter which is the most delicate and innermost layer of meningeal tissue surrounding components of the Central Nervous System.

5. the axon of an Alpha motor neuron connects with the extrafusal muscle fiber via neuromuscular junction which is a special type of chemical synapse. This chemical synapse differs in structure and function from chemical synapses that connect other type of neurons. As for example, although both synapses transduce electric signals to chemical signals using neurotransmitters, synapses of neuromuscular junction uses Acetylcholine while others use Glutamate or GABA.

6. Acetylcholine released from the neuromuscular junction during transmission of impulse from alpha motor neurons to extrafusal muscle fibers is sensed by nicotinic Acetylcholine receptors of the fibers which finally causes contraction in the muscles.

7. Alpha motor neurons have heavily myelinated axons for rapid information transmission in contrast to the axons of Gamma motor neurons.

Beta Motor Neuron

Beta motor neurons are responsible for innervating the intrafusal muscle fibers of the muscle spindles with collaterals to extrafusal muscle fibers [6]. In terms of functionality Beta motor neurons are generally of two types. They are:

1. **Slow Contracting:** These form of Beta neurons are responsible for innervating the extafusal muscle fibers.

2. **Fast Contracting:** These form of Beta motor neurons are respsonsible for innervating the intrafusal muscle fibers. Intrafusal muscle fibers are special type of skeletal muscle fibers that serve as specialized sensory organs known as Proprioceptor that detect the change of length in a muscle. They contain muscle spindles and are inverted by axons of one sensory and one motor neuron.

In terms of innervation of nuclear fibers of muscle spindles, Beta Neurons can be classified into two types. They are:

1. **Static Beta Neuron:** These Beta neurons innervate the nuclear chain fibers of the muscle spindles with collateral to extrafusal muscle fibers. Nuclear chain fiber is a special type of intrafusal muscle fiber or more precisely a sensory organ contained within the muscles which along with the Nuclear bag fiber is responsible for detection of change in length of a muscle. There are almost three to nine Nuclear chain fibers per muscle spindle and their nuclei are aligned in a chain that are responsible for exciting the secondary nerve. 'Nuclear chain' refers to the structure of central region of the muscle fiber where sensory axon wraps around the intrafusal fibers.

   The secondary Nerve association is involved in detecting stress and strain placed on a muscle which is later interpreted by the afferent and efferent pathways in order to measure the amount of stretch in a muscle and forwarded to the Central nervous System.

2. **Dynamic Beta neuron:** These Beta neurons innervate the nuclear bag fibers of the muscle spindles with collaterals to the extrafusal muscle fibers. The Nuclear bag fiber is a type of intrafusal muscle fiber that lies in the center of a muscle spindle. Each Nuclear bag fiber has a large number of concentrated nuclei in bags and causes excitation of both primary and secondary nerve. There are two types of Nuclear Bag fiber based on the speed of contraction and innervation of muscles. They are Bag1 and Bag2. These Bags are responsible for sensing dynamic length of a muscle and are sensitive to length as well as velocity.

Gamma Motor Neuron

Gamma Motor neurons are responsible for innervating the intrafusal muscle fibers and regulates the sensitivity of a muscle spindle to muscle stretching [7]. Intrafusal muscle contracts with the activation of Gamma motor neurons so that only a small stretch is required in order to activate the sensory neurons and stretch reflex. Gamma motor neurons can generally be divided into two types. They are:

1. **Dynamic Gamma Motor Neuron:** These motor neurons manages the BAG1 fiber of Nuclear bag fiber and enhances dynamic sensitivity.

2. **Static Gamma Motor Neuron:** These motor neurons manage BAG2 fibers of the Nuclear bag fiber and enhances stretch sensitivity.

Cell bodies of Gamma motor neurons are located in the anterior gray column of the spinal cord. They receive input from the reticular formation of Pons and Brain stem. They are not directly involved in contraction and expansion of muscles but helps in keeping the muscles taut so that alpha motor neurons can fire continuously leading to muscle contraction. It also plays a role in adjusting sensitivity of a muscle spindle.

Alpha Gamma Co-activation

When the Central Nervous System sends a signal to Alpha Motor neurons, it also gets transmitted to Gamma motor neurons in order to keep the muscle taut [110]. This is known as Alpha-Gamma co-activation. Without Gamma motor neuron, muscles would be very loose with the increase in contraction of the muscles. This would prevent the muscle spindles from detecting a precise amount of stretch.

Fusimotor System in Gamma Motor Neuron

It is a system by which the Central Nervous System controls the sensitivity of muscle. It consists of muscle spindle, Gamma and Beta neurons. As Beta Motor neurons innervate extrafusal and intrafusal muscles, it is known as Skeletofusimotor neuron. Gamma Motor Neurons are efferent part of the Fusimotor System whereas muscles spindles are afferent part as they send information from muscle towards brain and spinal cord.

Gamma Bias

The consistent level of activity of Gamma motor neuron is known as Gamma Bias. Smaller Gamma motor neurons require a small amount of excitatory input in order to reach the threshold. Therefore, Gamma motor neurons are likely to fire more than Alpha Motor neurons and hence creates a situation where Gamma motor neuron also fires consistently in the absence of a muscle stretch.

Visceral Motor Neuron

Visceral Motor Neurons control involuntary functions mediated by the activity of smooth muscle fibers, cardiac muscle fibers and glands [43]. Homeostatis is the continuous regulation of the expenditure and refilling of body's resources in order to maintain the balance of physiological functions of the body.Visceral Motor neurons are mainly controlled by the hypothalamus and the complex circuitry of the spinal cord.

General Visceral Motor Neuron

These Visceral motor neurons provide innervation to smooth muscles, cardiac muscles and glands of the body. The fibers of General Visceral Motor neurons can either be sympathetic or parasympathetic and exists in Oculomotor nerve, Facial nerve, Glossopharyngeal nerve and Vagus nerve of the Crania nerves.

Special Visceral Motor Neuron

Special Visceral Motor neurons provides innervation of the muscles of Pharyngeal arches in humans and Branchial arches in fish. It exists within Trigeminal nerve, Facial nerve, Glossopharyngeal nerve, Vagus nerve and Accessory nerve of the Cranial nerves.

Movement of Muscles due to Motor Neurons

A single motor neuron has the ability to innervate multiple muscle fibers and a single muscle fiber can undergo many action potentials in the time taken for a single twitch of the muscles. As a result the twitches can superimpose on each other if an action potential arrives before a twitch. This superimposition takes place either through summation of the twitches or through tetanic contraction.
In case of summation, the muscle is stimulated repeatedly in such a way that additional action potentials coming from the Somatic Nervous System arrive before the twitch. The twitches thus superimpose on each other generating a force greater than a single muscle twitch.
In case of tetanic Contraction, the action potentials arrive so rapidly that individual twitches are indistinguishable and tension rises smoothly eventually reaching a plateau.It is a form of constant but very high frequency stimulation of muscles.

2.1.3 Motor Neuron Disease

Motor Neuron Disease (MND) is a neurological condition that progressively damages the specialized nerve cells in our body known as motor neurons. This disease leads to progressive weakness of different muscles of our body that are involved in voluntary and involuntary movement. The effected regions generally are bulbar, thoracic, abdominal and limb muscles[22]. It has no known cause and is usually followed by death due to loss of muscle movement in organs that is used for ventilation like respiratory organs.

2.1.4 Types

Motor Neuron Disease can be of different types depending on the affected region of the Nervous System and origin. Some form of MND affects the upper motor neurons which is also known as spasticity or hyperreflexia while others can affect the lower motor neurons in the spinal cord which is also known as weakness or atrophy[16].

Amyotrophic Lateral Sclerosis (ALS)

This form of Motor Neuron Disease leads to Upper and Lower motor neuron degeneration. This is the most common type of MND and generally motor neuron disease refers to ALS. About eighty percent of patients of MND are affected by ALS. The symptoms of ALS tends to start from the muscles of hand and feet. The muscles initially seems to become stiff and weak which is accompanied by difficulty in swallowing and speech [70].

Hereditary Spastic Paraplegia (HSP)

This form of Motor Neuron Disease leads to degeneration of upper motor neurons only. In this form of MND the predominant symptom is confined to lower extremity (like the lower limbs) which may include weakening, involuntary spasm and muscle stiffness [51]. Other symptoms may also include gradual weakness in muscles, increased muscle tone, urinary problems, lack of sensation in feet, epilepsy, dementia, ichtyosis and loss of hearing.

Primary Lateral Sclerosis (PLS)

This form of Motor Neuron Disease (MND) leads to degeneration of upper motor neurons and pyramidal tracts in the brain and spinal cord with lower motor neurons preserved [32]. It is a rare form of Motor Neuron Disease which mainly causes weakness in legs and muscles. Some people is also known to have felt clumsiness in hand and speech.

Progressive Muscular Atrophy (PMA)

This form of Motor Neuron Disease leads to degeneration of lower motor neurons in the spinal cord. However, degeneration of upper motor neurons may occur in twenty to thirty percent of the patients with initial diagnosis within typically five years from onset up to 10 years [95]. The small muscles of the hand and feet are usually affected first but the muscles do not become stiff.

Progressive Bulbar Palsy (PBP)

It is a very rare form of Motor Neuron Disease which affects the Bulbar region of the lower motor neurons in spinal cord. The muscles involved in talking, chewing and swallowing i.e. the Bulbar muscles are affected in this form of MND.

Pseudobulbar Palsy

This form of motor neuron disease affects the Bulbar region of the upper motor neurons in the brain. The symptoms are slow and indistinct speech, difficulty in swallowing (also known as Dysphagia), Dysarthria, labile affect and brisk jaw jerk.

2.2 Machine Learning Algorithms

Learning is a many-faceted phenomenon that requires acquisition of new declarative knowledge, development of motor and cognitive skills through practice or instruction, generalization of acquired knowledge, effective representation of the knowledge and interpretation of new acts through observation and analysis [86]. The study of computer science in which algorithms are designed to provide machines the ability to learn in a similar way through experience and practice is known as machine learning. The process of Machine Learning is similar to that of data mining and predictive modeling. It requires searching through data in order to look for patterns and self adjust accordingly. Machine Learning can mainly be divided into two types. They are:

1. **Supervised Learning:** Supervised machine learning is a process in which the machines are trained first to perform a specific classification or regression using labeled and preprocessed data from which the machines try to find a pattern through trial and error that will be useful for predicting output from an unknown data [98]. This process is generally known as training the machine so that it can provide a desired output when supplied with an unknown data. Some examples of supervised machine learning algorithms are Support Vector Machine, Deep Neural networks, Random Forest Algorithm, etc.

2. **Unsupervised Learning:** Unsupervised Machine learning includes prediction of an outcome like supervised learning but it does not require to be trained from a set of known and labeled data first. As a result, although it does not have a means of calculating accuracy of the output it uses an iterative approach to extract various key features of the data in order to predict an output. Some of the examples of Unsupervised Machine Learning algorithms are K-Means clustering, Auto Encoder Neural Networks, Method of moment, etc.

2.2.1 Linear Regression

Regression is a statistical approach for detecting relationship between a variable of interest or prediction class and a set of predictor variables also known as features [80]. It is a method of supervised machine learning in which a set of training data is provided to a multidimensional feature space of some independent and one dependent variable. If multiple dependent variables exist that are not correlated to each other then the regression is known as Multiple Linear Regression. In case of multiple correlated variables, the regression is known as Multivariate Regression analysis. The model tries to find an equation that best fits the data in that given feature space. In simple linear regression this is done by finding a straight line that best fits the data and hence is known as the Best Fit Line.

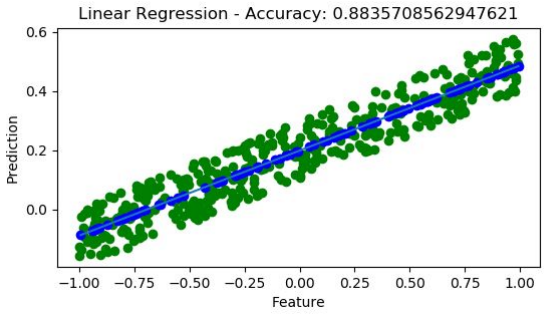

Figure 2.3: Linear Regression Graph

Figure 2.3 shows a graph of Linear regression where X-axis contains the independent variable also known as feature and Y-axis contains the dependent variable. The points in the graph are labeled data set that contains labeled prediction for the respective input features. The straight line is the best fit line that describes the data set in the feature space. Any value for the independent variable will produce the corresponding value for the dependent variable on the Best Fit Line.

The equation for the best fit line can be written as:

$$y = mx + b \tag{2.1}$$

Here 'm' is the slope, 'x' is the input value, 'b' is the y-intercept and 'y' is the predicted value for the given input.
The equation for determining the slope of Best Fit line is as follows:

$$m = \frac{\bar{X}\bar{Y} - \bar{XY}}{(\bar{X})^2 - \bar{X^2}} \tag{2.2}$$

Here $\bar{X}$ is the mean of all values of dependent variables of the training set and $\bar{Y}$ is the mean of all values of the labeled predicted or independent variable of the dataset. This equation is applicable for two dimensional feature space i.e. containing one dependent and one independent variable.
The equation for determining intercept of the best fit line is as follows:

$$b = \bar{Y} - m\bar{X} \tag{2.3}$$

The error for each data in the training dataset can be calculated using the R Squared algorithm by calculating error i.e. the square of the Euclidean distance between the data point and the best fit line. The equation for calculating the Co-efficient of determination is as follows:

$$Co - efficientofdetermination = 1 - \frac{R(\hat{Y})^2}{R(\bar{Y})^2} \tag{2.4}$$

Here $\hat{Y}$ is the best fit line. Co-efficient of determination is used to determine the quality of the best Fit line produced where high Co-efficient of determination indicates a better fitted line for the dataset. Hence, a best fit line has the highest Co-efficient of determination and thus can be used as a proof for the best regression. Accuracy of the model is directly proportion to the Co-efficient of determination.

2.2.2 K-Nearest Neighbor

K-Nearest Neighbor is a non-parametric supervised machine learning algorithm for classification where labeled data of different classes or groups are fed to a model which tries to group the data based on their respective classes [19]. As a result when an unknown data is fed to the model, 'K' nearest data points to the unknown data elects to determine to which class the unknown data point belongs. Here, 'K' is the user defined number of data points with the minimum Euclidean distance from unknown data. Normally the value of K is selected to be a number such that the election does not contain any tie. Generally it depends on the number of available classes.
The equation for the Euclidean distance between two data points is as follows:

$$Euclideandistance = \sqrt[2]{\sum_{i=1}^{n}(q_i - p_i)^2} \tag{2.5}$$

Here 'n' is the number of dimensions of features, 'q' is one point and 'p' is the other point whose distance needs to be calculated. Distance between unknown and each

data point is calculated this way and K nearest points with the minimum distance determines the class for unknown data.
One advantage of this algorithm is that the confidence of the model can be determined which is equal to the percentage of votes received for a correct predicted class.

2.2.3 Random Forest Algorithm

Random forest Algorithm is a supervised Ensemble Learning algorithm where multiple decision trees are constructed using labeled training data and a set of features for the data [108]. A set of such decision trees provide classification output for an unknown data and the final classification output is determined by the majority vote.

Decision Tree

It is an algorithm that takes a set of correlated data that shares some common features or attributes and splits the data into a logical tree in order to perform classification.The basic idea of Decision Tree is to build a tree greedily by choosing the 'most significant attribute' to be the root of the tree and then splitting the dataset based on the attribute values and repeating the process on new terminal leaves [62].

Day	Outlook	Humidity	Wind	Play
1	Sunny	High	Weak	No
2	Sunny	High	Strong	No
3	Overcast	High	Weak	Yes
4	Rain	High	Weak	Yes
5	Rain	Normal	Weak	Yes
6	Rain	Normal	Strong	No
7	Overcast	Normal	Strong	Yes
8	Sunny	High	Weak	No

Figure 2.4: Example of Decision Tree

Figure 2.4 shows a table containing of 8 inputs data each of which is a row containing the weather, wind and humidity condition for each day and a decision based on the three features whether 'X' played on that day.
Therefore, the decision tree for the prediction of whether 'X' will play if the outlook of the weather is 'Rain', Humidity is 'High' and the wind is 'Weak':
Figure 2.5 shows a decision tree for whether 'X' will play in the provided conditions or not. The shaded region indicates nodes of the tree that are features of the dataset and are selected using algorithms like Information Gain, GIGI index approach, etc. The leaves without any child are 'Pure subsets' as the node provides only one decision output. The non-shaded circles indicate all possible values of the attribute. The attribute 'Vote' indicates the number of votes provided by the specific node for each output. It is also known as the 'Confidence'.
In order to determine whether X will play if the outlook of the weather is 'Rain', Humidity is 'High' and the wind is 'Weak' can be determined by traversing the decision tree in the following way:

1. Start from root node 'Outlook'. For the problem, as the value of outlook is 'Rain', traversing down that parameter provides the feature Node 'Humidity'.

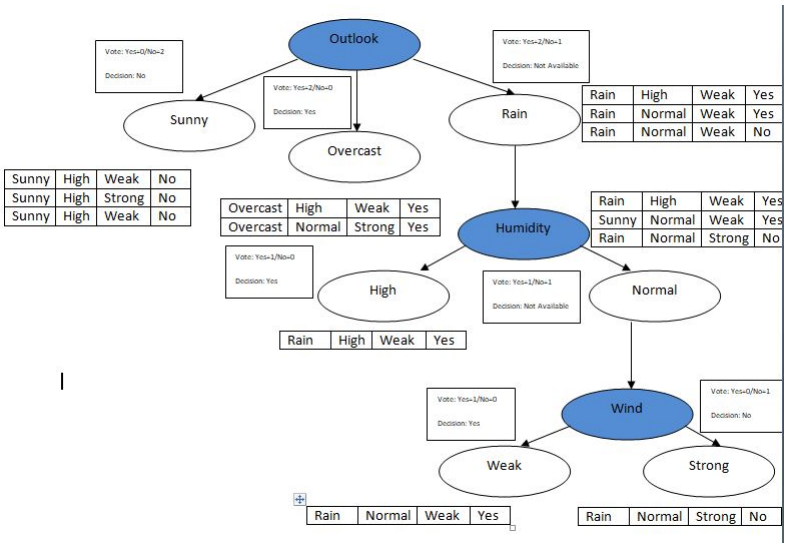

Figure 2.5: Example of Decision Tree(contd.)

2. Traversing down the node of 'Humidity' the node 'High' is chosen to be traversed next as it is the value of the required Humidity.

3. Since the node 'High' is a pure subset, the decision of this node is taken as the final decision value which is 'Yes'.

4. Therefore according to the classification of the decision tree, X will play in the aforementioned conditions.

Iterative Dichotomizer(ID) 3 Algorithm

ID3 is an iterative approach invented by Ross QUinlan to generate a decision tree [11]. It splits a set of training data among each leaf node in order to create a set of pure subsets that will take part in decision making process. The algorithm is as follows:
$-split(node, dataset)$:

1. A $<-$ The best attribute for splitting dataset

2. attribute $<-$ Decision tree attribute for node A

3. For each value of A create a child node and add it to a list

4. Split dataset among the child nodes.

5. For each child node in the list, if the subset is not pure then $split(childnode, subsetofdatasetinth$

Choosing the best attribute

In order to choose the best attribute from the list of features in the data set we need to expand(or split into child nodes) each feature and measure purity for each split so that the machine can decide which split will be the most efficient. The metric used for measuring purity of the split is known as 'Certainty'.

Entropy

It is the mechanism to measure uncertainty of a class in a subset of examples. It provides a value between 0 and 1 which indicates the rate of uncertainty regarding a certain decision. The equation for Entropy is:

$$H(S) = -P_{+}.log_2(P_{+}) - P_{-}.log_2(P_{-}) \tag{2.6}$$

Here 'S' is the subset of training examples, P_{+} and P_{-} are the percentage of positive and negative examples in the set. $H(S)$ is the entropy or uncertainty for the training subset.

Information Gain

Information Gain is an algorithm that aggregates information from different child nodes of a decision tree in order to measure purity or uncertainty of a parent node [96]. Generally, the number of pure(positive or negative) subsets provides a better purity for the parent node. The equation for the expected drop in entropy after splitting training examples among the child node is as follows:

$$Gain(S, A) = H(S) = \sum_{v \epsilon Values(A)} \frac{|S_v|}{|S|} H(S_v) \tag{2.7}$$

Here 'v' is the possible values for attributes 'A', 'S' is the set of training examples 'x' and 'S_v' is the subset of examples where $x_A = v$. $Gain(S, A)$ is the average purity for the attribute node. An attribute with the highest gain has the highest priority of being chosen as the attribute node to be expanded in the decision tree.
Some of the problems with Information Gain is that it is biased towards the node having many values whether the values be the one which will provide the desired output or not. Moreover, the algorithm does not work if a new value for a specific feature or attribute of the training dataset is added. In this case, the best solution is to use Gain Ratio whose equation is as follows:

$$GainRatio(S, A) = \frac{Gain(S, A)}{SplitEntropy(S, A)} \tag{2.8}$$

Here Gain Ratio penalizes the attribute with many values in order to reduce the bias. The equation for $SplitEntropy(S, A)$ is as follows:

$$SplitEntropy(S, A) = -\sum_{v \epsilon Values(A)} \frac{|S_v|}{|S|} .log \frac{|S_v|}{|S|} \tag{2.9}$$

Here 'A is the candidate for Attribute whose Gain Ratio has to be calculated, 'v' is the possible values of 'A', 'S' is the set of training examples 'X' and 'S_v' is the subset where $X_A = V$

Multi Class Classification

The equations provided are for a decision tree that can perform binary classification(0 or 1). But for generating decision trees that can perform Multi Class Classification the structure of the algorithm will be the same but the tree will predict most frequent class in the subset of nodes. The equation for Entropy will also change in this case.

Regression

A Decision tree is also able to perform Regression where the predicted output will be the average of training examples in the subset. It will also require a different equation for entropy and can use linear regression at leaves.

Random Forest

It is an algorithm where multiple decision trees are generated and each tree performs independent classifications for a specific prediction and the final output will be selected by the votes of the trees where each tree votes for a specific output.

2.2.4 Support Vector Machine

Support Vector Machine is a supervised machine learning algorithm invented by Vladimir N. Vapnik and Alexey Ya. Chervonenkis in 1963 [23]. It can perform both classification and regression by creating a Maximum Margin Hyperplane that lies in a transformed input space and splits the example classes while maximizing th distance to the nearest cleanly split example [60].

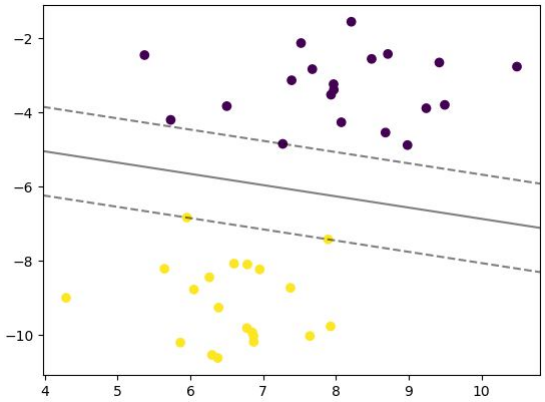

Figure 2.6: Support Vector Machine Graph

Figure 2.6 shows the output graph for a Support Vector Classification where the points represent the coordinates of training data in a two dimensional feature space and color of the point represents a class e.g. yellow indicates data points of class '0' and purple points indicate the class '1'. The dashed(–) straight lines represent the edges of the hyperplane separating data points of the two classes. The points on the dashed lines are known as support vector, the line in between the dashed line represents the median of the hyperplane and the dashed rectangular region is the Maximum Margin Hyperplane.

Maximum Margin Classifier

It is a hypothetical classifier where the number of input space forms an 'n' dimensional space where 'n' is the number of input variables.

Hyperplane

It is a line in Maximum Margin Classifier that splits the input variable space. In Support Vector Machines, the hyperplane separates the input space based on specified class.

Margin

It is the perpendicular distance between hyperplane and closest data points to it. The data points are known as support vectors.

Maximum Width

Distance or width between the support vectors of the split classes in an input space is known as the 'Street'. Width of the split is directly proportional to the efficiency of the model. So, the maximum width of the street can be calculated as follows:

$$L = \sum_i \alpha_i - \frac{1}{2} \sum_i \sum_j \alpha_i.\alpha_j.y_i.y_j.\vec{x_i}.\vec{x_j} \tag{2.10}$$

Here, α represents Langragian multipliers, 'y' represents the classification output for training data and 'x' represents input training data vector. 'L' is the maximum width of the hyperplane due to constraints $y_i(\vec{w}\vec{x_i} + b) - 1$.

2.2.5 Neural Networks

Artificial Neural Networks can be characterized as computational models inspired from biological neurons and neural network with particular properties such as the ability to learn and adapt, to generalize, cluster or to organize data. The first computational model for neural networks was designed by Warren McCulloch and Walter Pitts in 1943 [2]. However, as we still know a very little about biological system, the models designed are oversimplification of biological neural networks [29]. An Artificial Neural Network(ANN) is composed of a large number of highly interconnected processing elements(also known as neuron or perceptron) working in unison to solve computational problems like pattern recognition or data classification through adaptive learning.It is a form of supervised machine learning.

Types

There are different kinds of Neural Networks designed using different computational models starting from a single cell neuron to complex multi-layered neural networks with memory.
Figure 2.7 displays different models of Artificial Neural Networks used for different kind of computational problems.

1. **Perceptron:** Perceptron is the fundamental building block of an Artificial Neural Network just like a neuron which is the fundamental building block of a biological neural network. However, unlike biological neurons which process information using electrical and chemical medium, Perceptrons use numerical data (also known as weights and biases) and mathematical equations in order to process information. From the viewpoint of Computer Science, it is an

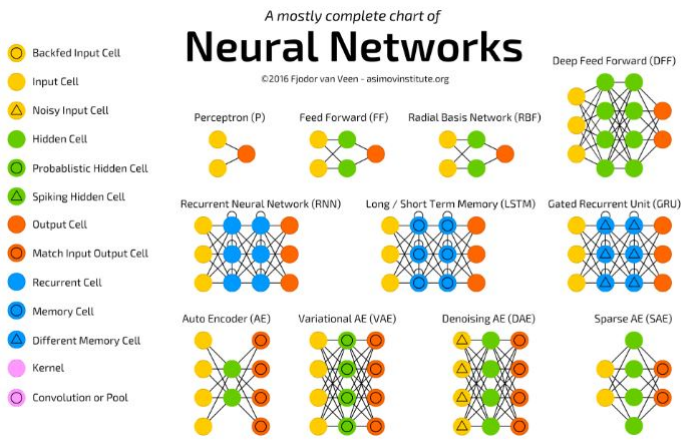

Figure 2.7: Architectures of Neural Network

algorithm for supervised machine learning that can perform binary classification for linearly separable problems. This model of Neural Network was first designed by Frank Rosenblatt in 1957 at the Cornell Aeronautical Laboratory [4].

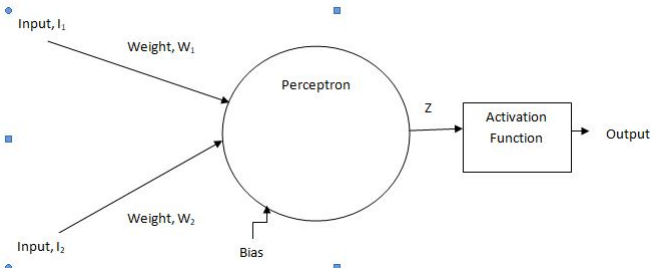

Figure 2.8: Structure of a Perceptron

Figure 2.8 represents the structure of a Perceptron with two inputs and one output. Each input connection contains an adjacent weight which is an adjustable numerical value which is multiplied by their respective inputs. The sum of these weighted inputs determine the output when passed through an activation function. Each perceptron has a bias which is also an adjustable numerical value used in order to remove ambiguity from the predicted output. In this way a single perceptron performs binary classification for a set of inputs and this process of prediction is also known as forward propagation.

$$Z = \sigma(\sum_{i} W_i I_i + B) \tag{2.11}$$

Here 'W' represents the weight for Input 'I', 'σ' represents the activation function and 'B' represents the bias of the Perceptron for predicting an Output 'Z'.

2. **Feed Forward Neural Networks:** Neural Networks where the data flow from input to output unit is strictly feed forward and can exist over multiple layers each containing multiple perceptrons without any feedback connection are known as Feed Forward Neural Network.

3. **Radial Basis Network:** Radial Basis Networks are multi-layered Neural Networks containing Radial Functions which is a special type of function whose characteristic feature is that the response changes monotonically with distance from a center point. It was first formulated by Broomhead and Lowe in 1988 [12]. The center, distance from the center and precise shape of the radial functions are parameters to the model. A Radial Basis Network is linear for fixed parameters and non-linear if the radial functions can move, change size or contains multiple hidden layers. There are different kinds of radial function. The equation for a Gaussian Radial Function for scalar input is as follows:

$$h(x) = exp(-\frac{(x-c)^2}{r^2}) \tag{2.12}$$

   Here 'c' is the center, 'r' is the radius and 'x' is the input to Gaussian Radial Function.

4. **Deep Feed Forward Network:** Deep Forward Neural Networks are multi-layered Feed Forward Neural Network that can perform classification for multiple output classes. This form of Neural Network can perform classification for linearly inseparable problems. Convolutional Neural Networks(CNN) are a form of Deep Feed Forward Neural Networks.

5. **Recurrent Neural Networks:** Recurrent Neural Networks are a form of Artificial Neural Network designed in 1980s for pattern recognition taking time and sequence into account i.e it possesses a temporal dimension. Hopfield networks were discovered by John Hopfield in 1982. Each hidden unit in the network contains a form of temporary memory where an output is stored and used with the next input for classification. This creates a relationship between each sequences of input and hence provides a temporal dimension.

6. **Long/Short Term Memory Network:** Long/Short Term Memory(LSTM) Networks are a form of Recurrent Neural Network that is capable of learning long-term dependencies. It was discovered by Hochreiter and Schmidhuber in 1997 [30]. Its default behavior is to remember information for long period of time. Even though LSTM has a chain like structure for remembering information, unlike RNN it has a different structure for repeating modules. The cells in LSTM has the ability to add or remove information from cell states which is carefully regulated by structures called gates. Gates are way to regulate the transmission of information by either permitting or denying the passage of an information.

7. **Gated Recurrent Unit Network:** Gated Recurrent Unit(GRU) Networks are a form of RNN that solves the Vanishing Gradient Problem of RNNs by using gates like LSTM Networks. It was first introduced in 2014 by Kyunghyun Cho et al [92]. In order to solve the Vanishing Gradient Problem GRU uses update and reset gates with Long Term Memory.

8. **Auto Encoder Networks:** An Auto Encoder(AE) Neural Network is a type of Artificial Neural Network which is used to learn efficient data encoding using unsupervised machine learning mechanism. It encodes a set of data

generally for the purpose of dimensionality reduction and is often integrated within sparse and deep neural networks. Some of the forms of Auto Encoders are Variational AE that can make strong assumption about the distribution of latent variables, Sparse AE which solved the difficulties which were associated with training traditional AE and Denoising AE that takes partially corrupted input and while training in order to recover the original undistorted input.

Activation Functions

In Artificial Neural Networks inspired from Biological Neural Networks, the activation function is usually an abstraction representing the rate of action potential firing in each neuron or perceptron. In the most simplest form the function is binary i.e. it decides whether the perceptron will fire or not. This form of activation function is known as Threshold functions. There are other forms of activation function e.g. identity function, binary step function, bipolar step function, sigmoidal function, ramp function, etc. Some of the popular non-linear activation functions are:

1. Rectified Linear Unit(ReLU) function [Range: 0 to ∞]
2. Tanh function [Range: -1 to 1]
3. Sigmoid function [Range: 0 to 1]
4. Leaky ReLU [Range: -∞ to ∞]

Some of the desirable properties of activation functions are:

1. The function should be non-linear.
2. The function should be continuously differentiable.
3. The range of activation function should be finite as it makes gradient based optimization methods more stable.
4. The function should be monotonic or should have a monotonic derivative.

Cost Function

It is a function that evaluates whether the current set of weights in an Artificial Neural Network are doing a good job for regression. In other words, it is used to determine the amount of loss for a predicted output while training an ANN which is later used to optimize the network. As a result it is also known as Loss Function. One of the most common form of Cost Function used is Mean-Squared Error(MSE) loss Function. The equation for MSE function is as follows:

$$J(\theta_i) = \frac{1}{2m} \sum_{i=1}^{m} [h_{\theta_i}(x^{(i)} - y^{(i)})]^2 \tag{2.13}$$

Here 'm' is the number of items in the training dataset, 'i' is the i-th element of the input training dataset and 'y' is the label or desired output for the i-th element of the dataset.

The more generalized form of equation for calculating the Cost function for a particular training dataset is as follows:

$$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}\sum_{k=1}^{K}[y_k^i log((h_\theta(x^i))_k)+(1-y_k^i)log(1-(h_\theta(x^i))_k)]+\frac{\lambda}{2m}\sum_{l=1}^{L-1}\sum_{s=1}^{S_L}\sum_{j=1}^{S_L+1}(\theta_{j,s}^l)^2 \tag{2.14}$$

Here 'm' is the number of input dataset, 'L' is the number of Layers in the network, 'K' is the number of output classes, 'S_L' is the number of neurons in Layer 'L', 'θ^l' is the weight matrix of layer 'l', 'y_k' is the desired output for class 'k' and '$h_\theta(x^i)$' is the predicted output for input data 'x^i'.

2.3 Diagnosis of Motor Neuron Disease

In order to perform diagnosis of MND data can be extracted in different formats from the patients. Some of the data acquisition techniques are discussed below.

2.3.1 ALS Functional Rating Scale

One of the data acquisition techniques contain questionnaire-based scale that measures the physical functioning while the patient carries out daily activities of living. This form of data acquisition technique is known as ALS Functional Rating Scale(ALSFR) [104]. However, this technique is useful for tracking the progress of the disease rather than detection of it as the technique relies on verbal response of patients and can only be used when the disease progresses so far that the patient starts to notice physical problems.

2.3.2 Medical reports

Another form of data acquisition technique requires the patient to undergo different medical tests which is used for the detection of ALS. This dataset contains clinical reports, lab reports, demographic data, family history, etc. of a patient which is used to detect the disease.Pooled Resource Open-Access ALS Clinical Trials Database(PRO-ACT) is an example of a web application that contains dataset of such kind. These reports help to detect various symptoms of ALS in a patient and also helps to rule out other diseases that might mimic ALS.

2.3.3 Nerve Conduction Study

Nerve Conduction Study(NCS) is a medical diagnostic test that evaluates the ability of electrical conduction of the motor and sensory nerves of the human body by producing artificial stimulus to the nerves and measuring the response in different regions of the corresponding motor units. It is an integral part for the evaluation of ALS in patients and is also useful in distinguishing ALS from other neuropathic and myopathic diseases. It is usually done by measuring whether abnormal conduction velocity, distal latency, F-wave latency, etc. exceeds from what is expected from ALS alone [20].

2.3.4 Brain Computer Interface

Brain Computer Interface(BCI) is a communication pathway between a wired brain and an external device. It provides pathway for bidirectional communication with the brain and is often used for researching, mapping, assisting, augmenting, or repairing human cognitive or sensory-motor functions [99]. This technique uses signals recorded from the scalp, surface of the cortex or from inside the brain in order to communicate with the external devices. Electrodiagnostic devices like Electroencephalography(EEG) and Surface EMG are generally used for this purpose. Even though it is a comparatively new technology researches are currently being performed to use BCI for analyzing different neuropathic diseases like ALS [53].

2.3.5 Electromyography

It is an electrodiagnostic technique which is used for evaluating and recording the electrical activity produced by different skeletal muscles of the body [89]. It is done with the help of an electromyograph and shows the waveform of the recorded electrical activity as a function of time and amplitude. The amplitude generally represents the electrical voltage produced at in instant in time which might correspond to force generated by the muscle under examination. In case of isometric contraction of muscles, the relationship between muscle force and smooth rectified Electromyograph(EMG) signal is usually linear [31]. As ALS is a progressive neurodegenerative disease that gradually destroys the lower motor neurons of the body and affects the electrical current of the muscles. EMG is a very useful device for detection of ALS as it can directly measure the electrical current of different regions of the body. However, raw EMG signals contain a lot of noise information (also known as artifacts) because of external factors like ambient noise, inherent noise in electronics equipment, motion artifact, muscle cross talk, baseline shift, nearby electrical equipment like cell phones or lights, etc. [103]. As are result the EMG signal needs to be filtered, smoothed and preprocessed before use in order to obtain important information that might be useful for classification of different diseases like ALS.
EMGLab is an example of a database containing EMG signal records of different neuropathic(ALS), myopathic and normal subjects recorded using different types of Electromyograph (like monopolar, concentric, quadrifilar, etc. needle electrodes) and different types of muscular contraction (like isometric, ramp and trapezoidal contraction) [41].

2.4 Electromyography and ALS Classification

Due to its complex nature, a lot of experiments have been conducted in order to extract information from EMG Signal. Some of the unconventional feature extraction techniques include Linear Predictive Coding(LPC), Multiscale entropy-based approach, Ensemble-Empirical-Mode-Decomposition-Based Intrinsic Component Analysis(ICA), second order volterra series, etc[34], [63], [82], [101]. However, state-of-the-art techniques used for feature extraction from EMG signals can be generalized into three primary types: time domain analysis, frequency or time-frequency domain analysis and feature extraction from decomposed EMG signals[18], [26], [38],

[56], [61], [77], [90], [91], [107]. Fattah et al.[78] used Time and Frequency Domain Features along with Fast Fourier Transform in order to classify healthy subjects and subjects affected by Amyotrophic Lateral Sclerosis. Pandey et al. used integrated intelligent computing model for the interpretation of EMG based neuromuscular diseases[59]. Christodoulou and Pattichis[36] used Self-organizing Feature Map(SOFM) and Learning Vector Quantization(LVQ2) technique in order to decompose EMG signals into potential MUAPs and to extract different features from the MUAP waveform and their firing rates. They conducted the experiment using 40 subjects and acquired a success rate of 97.6% and 95.3% when EMG decomposition was performed using Artificial Neural Networks and statistical learning respectively. Subasi used soft computing techniques, multi-layer perceptron neural network, dynamic fuzzy neural network and adaptive neuro-fuzzy inference system in order to perform classification of EMG signals using different feature extraction method like Wavelet Transform and Frequency Domain Features[84]. Rafiee et al. used wavelet domain features in order to classify EMG signals[73].

Chapter 3

System Architecture

This chapter contains a detailed description about the insights and procedures followed in order to collect and process EMG signal data for the diagnosis of ALS in Section 3.1. It also discusses the system architectures developed for classification of ALS using Time-Frequency Analysis[78], Discrete Wavelet Transform[90] and EMG Decomposition[36] in Section 3.2. These classifications were performed in order to compare the change of performance of these classification algorithms as mentioned in Section 1.3.1. Section 3.3 discusses about a new system architecture developed for the classification of ALS using selected features from multiple domain representation of the data.

3.1 Data Collection and Preprocessing

3.1.1 Data Type

The total dataset consists of three main attributes:

1. Subject type: The data acquired are either collected from or simulated for healthy subjects and subjects affected via ALS.

2. Repository: The data can either be from repository of Nikolic M. at Faculty of Health Science, University of Copenhagen in 2001[41] or from the repository of Hamilton-Wright[48].

3. Signal Type: The data can either be a real EMG signal acquired from healthy and neuropathic subject or it can be a simulated EMG signal representing the same muscle location.

3.1.2 Subject Group(Nikolic M.)

The material consisted of a normal control group and a group of patients with ALS. The control group consisted of 10 normal subjects aged 21-37 years of which 4 of them are females and 6 of them are males. None in the control group had signs or history of neuromuscular disorders. The ALS group consisted of 8 patients of which 4 are female and 4 of them are male aged 35-67 years. Besides clinical and electrophysiological signs compatible with ALS, 5 of the them died within a few years after onset of the disorder, supporting the diagnosis of ALS[41].

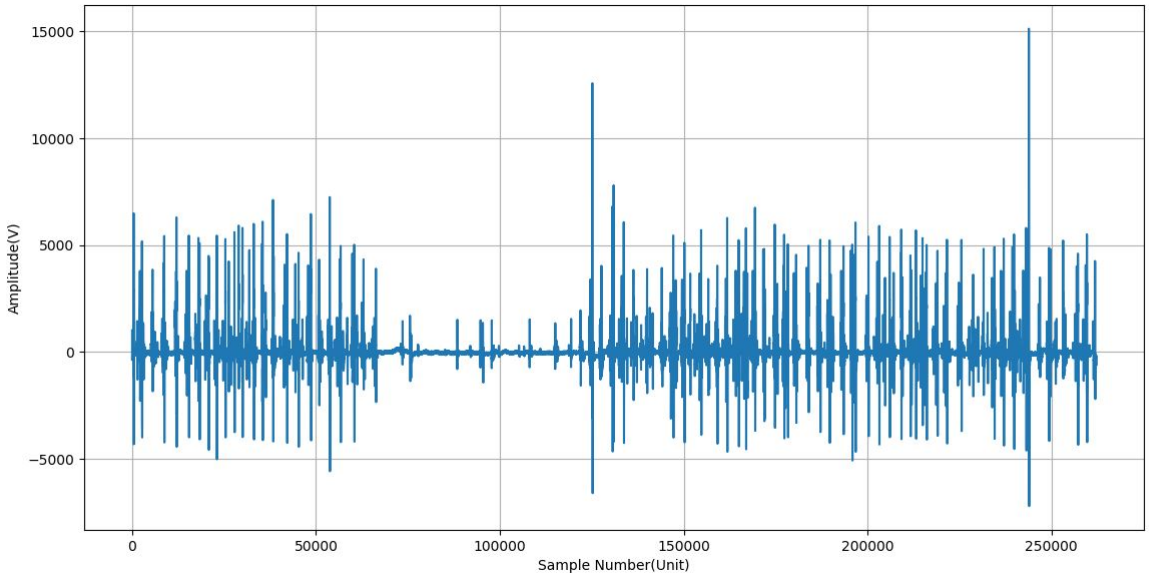

Figure 3.1: EMG signal acquired from a subject suffering from ALS - by Nikolic M.

Figure 3.1 and 3.2 shows the EMG signals acquired from a subject suffering from ALS and a normal subject respectively. The data was acquired from the repository of Nikolic M. X-axis represents the sample points in the signal and Y-axis represents the amplitude(uV).
The EMG recording conditions for the dataset are provided below:

1. The recordings were made at low voluntary and constant level of contraction.
2. Visual and audio feedback was used to monitor the signal quality.
3. A standard concentric needle electrode was used.
4. The EMG signals were recorded from five places in the muscle at three levels of insertion (deep, medium, low).
5. The high and low pass filters of the EMG amplifier were set at 2 Hz and 10 kHz.

Table 3.1 shows the description of each patient from whom EMG recordings were acquired. The ID column show each subject ID followed by his/her age, gender, status(whether the subject is healthy or suffering from ALS), duration(the duration of progression of the disease) and the number of EMG signals recorded from the subject.

3.1.3 Subject Group(Hamilton-Wright)

The data are simulated EMG signals generated for Normal and Neuropathic subjects using the simulator published as [48]. Myopathic/Neuropathic data has been simulated for 25, 50 and 75% fibre/motor unit involvement. Each study contains 5 contractions, each is in the range 7.5-12.5 MVC. EMG signals demonstrate the muscle activity of Biceps Brachii.

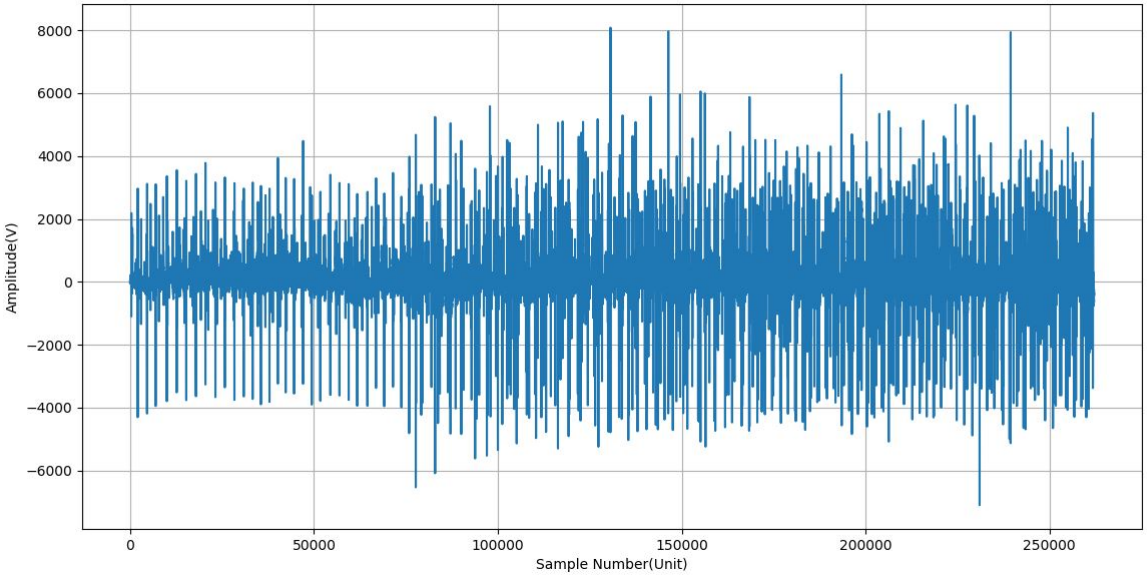

Figure 3.2: EMG signal acquired from a healthy subject - by Nikolic M.

SL No.	ID	Age	Gender	Diagnosis	Duration	Total
1	a01	56	M	ALS	0.5 years	12
2	a03	61	F	ALS	1.5 years	20
3	a04	67	F	ALS	0.5 years	20
4	a05	52	F	ALS	1.5 years	3
5	a06	56	M	ALS	0.5 years	17
6	a07	65	M	ALS	1 year	10
7	a08	60	F	ALS	¡0.5 years	15
8	c01	29	M	Normal	N/A	27
9	c02	26	F	Normal	N/A	23
10	c03	37	M	Normal	N/A	30

Table 3.1: Details of subjects - from the repository of Nikolic M.

3.1.4 Data Format/Structure

1. The main dataset directory consists of two folders (train and test) where the 'test' folder contains patient records for testing the classifier and 'train' folder contains patient record for training the classifier.

2. Each of the train/test directory consists of two folders(ALS and normal) where each folder contains folders for different subjects who falls under the specified group.

3. The record folders of each subject is stored in a folder(patient folder) bearing a unique ID number(e.g. `a01_patient`, `c01_patient`, etc.) for each individual patient. Each patient folder can have multiple EMG record folders obtained from the brachial biceps of the subject. Each record folder contains information related to each signal recorded from the specific patient.

4. Each record folder of the subjects also bear a unique ID number(e.g. N2001A01BB05,

N2001A01BB06, etc.) and each folder contains three files. They are 'data.npy' and 'data.hea'.

5. **data.npy:** This file contains the EMG signal recorded from an electromyograph. The data is stored as a Numpy one dimensional array where the length of array indicates the number of samples obtained at a specific sampling frequency. Number of samples for each signal data is 262124 for a sampling frequency of 23437.5Hz.

6. **data.hea:** This is a header file that contains all the information regarding subject under investigation and recorded EMG signal from the subject. As for example it contains the sampling frequency and total number of samples obtained from the signal, gender of the subject, period of diagnosis, duration of disease, location of placement of electrode, filters used, level of insertion of needle, etc. The data is stored as a text file.

3.1.5 Data Preprocessing

The raw digital signal obtained from an electromyograph is not very useful as it contains a lot of noise in the data i.e. it has a very low signal to noise ratio. It is due to the fact that the recording of nerve activity of a specific muscle region is interfered due to external factors like nearby electrical equipment. The voltage reading obtained while recording is often a mixture of voltages obtained from multiple nearby electrical sources [105]. Since there might be multiple cause of random noise, its signal to noise ratio cannot be known a priori [66]. As a result raw EMG signal needs to be preprocessed first in order to filter out the noises as much as possible and keep the important information for further processing and classification. The Power Density Function of Surface EMG signals contain most of the information in between 5-10Hz to 400-450Hz. So, these signals are usually filtered within a range of 5Hz(High Pass) to 500Hz(Low Pass). On the other hand, as intramuscular and needle EMG is recorded from within the muscles, it is usually filtered with a minimum low pass filter of 1500Hz [37]. Moreover, one of the most important source of noise for signals with typically low amplitude(mV) is the interference caused by the power lines(50/60Hz) and their harmonics[42]. Usually these type of noises are removed using a notch filter. The steps included for preprocessing the data are discussed below. The signal in the experiment was filtered using a second order Butterworth Filter with a low pass frequency of 1500 Hz and a High pass frequency of 20Hz.

3.2 System Architecture for "Review of different EMG feature extraction and classification techniques and their limitations during the development of a general purpose application software"

Time-Frequency Analysis(TFA), Discrete Wavelet Transform(DWT) and EMG Decomposition Algorithm(EDA) are three of the commonly used techniques for the

classification of Electromyography(EMG) signals of neuropathic, healthy and myopathic subjects. However, the performances of these techniques vary with respect to the dataset using which the classification is being performed. As for example, our study shows that the accuracy of TFA ranges around 70-100% when conducted under some experimental setup whereas its accuracy drops down to a range of 50-70% and 70-90% for real clinical and simulated EMG signal data when the classification is performed using open data. Furthermore, the accuracy of DWT fluctuates around 95-97% for some experimental data whereas it fluctuates around 60-80% and 70-75% for real clinical and simulated EMG signals respectively when conducted with open data. Similarly, the accuracy of EDA ranges around 60-70% and 70-75% for real and simulated signals respectively. Such variation of performance acts as a barrier to the development of an automated and adaptable software capable of consistent performance on a wide variety of dataset.
The system architecture developed performs a systematic comparison of Time-Frequency Analysis(TFA), Discrete Wavelet Transform(DWT) and EMG Decomposition Algorithm(EDA) for classification of ALS and healthy subjects based on the variation of dataset. Classification was performed for features extracted from each of the three techniques according to the procedure mentioned in [78], [90] and [36]. The variation of dataset are based on three factors: change of data repository which shows the variation of performance for dataset acquired from two different clinical environments, change of size of input data for training a classifier which shows the adaptation capability of the algorithm to varying size of training data and change of signal type which shows the variation of performance for dataset acquired for two different types of signal like simulated EMG signals and Real EMG signals.

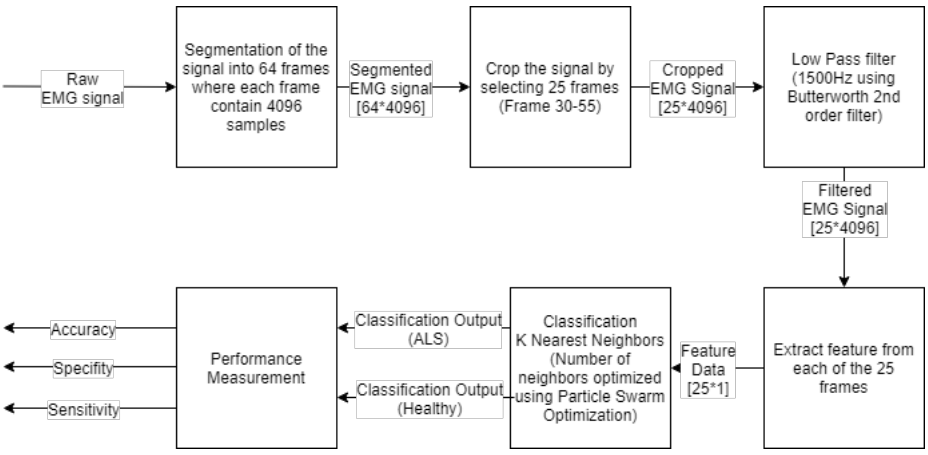

Figure 3.3: Block Diagram for classification using TFA

Figure 3.3, 3.4 and 3.5 shows the block diagrams of the system architecture used for the classification of ALS using TFA, DWT and EDA respectively. Raw EMG signals are provided as input to the system and a metric for the measure of performance of the classifier along with predicted class is obtained as the final outputs. The complete algorithms are provided in [78], [90] and [36].

3.2.1 Dataset

The total dataset consists of three main attributes:

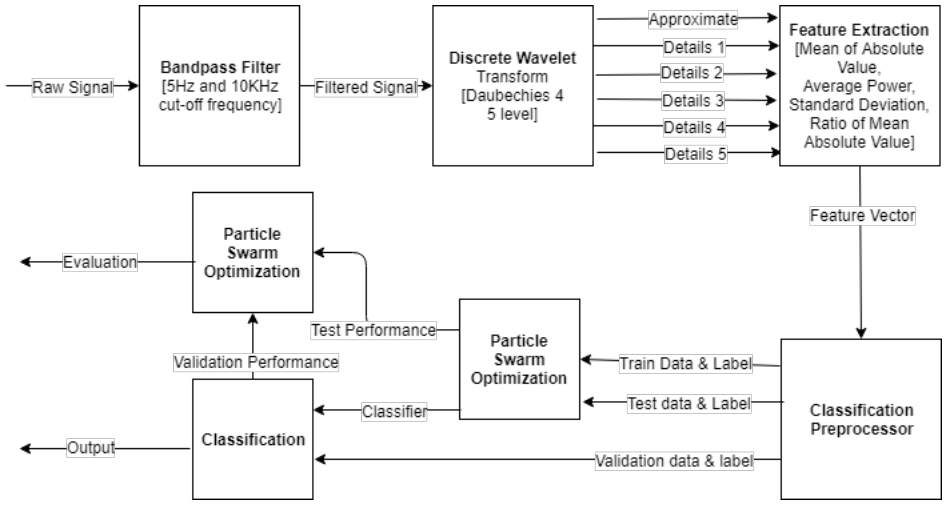

Figure 3.4: Block Diagram for classification using DWT

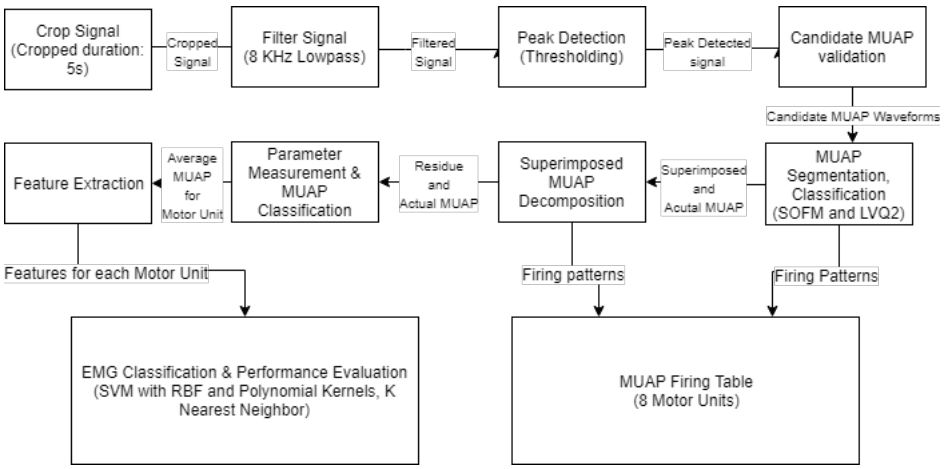

Figure 3.5: Block Diagram for classification using EDA

1. Subject type: The data acquired are either collected from or simulated for healthy subjects and subjects affected via ALS.

2. Repository: The data can either be from repository of Nikolic M. at Faculty of Health Science, University of Copenhagen in 2001[41] or from the repository of Hamilton-Wright[48].

3. Signal Type: The data can either be a real EMG signal acquired from healthy and neuropathic subject or it can be a simulated EMG signal representing the same muscle location.

Data from other repositories

The main purpose of this research is to compare the classification performance for a specific classification technique based on variation of input dataset. The classification techniques which will be applied for analysis are:

1. Identifying the motor neuron disease in EMG Signal using Time and Frequency Domain Features with comparison by Fattah et al.[78].

2. Classification of EMG signals using PSO optimized SVM for diagnosis of neuromuscular disorders by Subasi[90].

3. Unsupervised Pattern Recognition for the Classification of EMG Signals by Christodoulos et al.[36].

3.2.2 Feature Extraction

The following features were extracted for each of the classification techniques:

Feature Extraction using Time-Frequency Analysis

In the paper presented by Fattah et al.[78] data was extracted from the time domain of low-pass filtered data and frequency domain after performing Fourier Analysis of the data.The features that were used in this study for the classification of ALS included 'Average Amplitude of Spectral Peaks(ASP)' and 'Mean Frequency(MNF)' obtained from magnitude spectrum of FFT of segmented data, 'zero lag of Autocorrelation(ZLA)' and 'Zero Crossing rate(ZCR)' of time domain of the segmented EMG signal. Classification was performed based on each of these features extracted from each segment(frame) of a segmented EMG signal.

Feature Extraction using Discrete Wavelet Transform

In the paper presented by Subasi[90], data was extracted MAV, AP and SD of the detail(Level 1 to Level 5) and approximate(Level 1) coefficients of DWT in each sub-band for a 5th level decomposition of EMG using db4 wavelet.

Feature Extraction using MUAP Decomposition

In the paper presented by Christodoulos et al.[36], features were extracted from the MUAP waveform and firing pattern identified after decomposing a raw EMG signal to its constituent potential MUAPs using EDA. The features used for classification included 'Amplitude difference(AD)' between maximum negative and minimum positive peaks, 'Duration' from the beginning of the MUAP waveform where the signal is greater than a threshold equal to 1/15 of the amplitude for the MUAP waveform, Area of Rectified MUAP(ARM) integrated over the calculated duration, Rise Time(RT) between maximum negative peak and the preceding minimum positive peak within the duration, 'Phases' or Number of baseline crossings within the duration where amplitude exceeds 25 V(plus one) and 'Number of Turns(NOT)' or positive and negative peaks where the differences from the preceding and following turn exceed 25 V for the MUAP waveform.

3.2.3 Classification

The following three classification techniques were used for the three domain features extracted in previous step:

Classification for Time-Frequency Analysis

KNN was used to classify subjects with ALS and healthy subjects[78]. The number of neighbors used in the classification was not mentioned in the paper and hence PSO was used to calculate the parameter for the number of nearest neighbors.

Classification for Discrete Wavelet Transform

The classifiers used for the classification of EMG signal are KNN, SVM and SVM optimized using PSO[90]. Random Forest Algorithm(RFA) is also used alongside these classifiers in order to analyze the performance of RFA in classifying EMG using the features extracted from DWT.

Classification for MUAP analysis

The paper presented by Christodoulos et al.[36] SOFM and SOFM with LVQ2 was used to classify healthy subjects and subject affected by neuropathy.

3.2.4 Comparison strategy

In order to analyze the performance of different classification techniques with varying dataset the following steps were followed:

1. The signals were preprocessed(filtered, cropped, etc.) according to format defined in the respective papers[78][90][36].

2. The features were extracted according to format defined in the papers.

3. Classification was performed using classifiers and parameters mentioned in the papers. In case of absence of mention of a classifier parameter, PSO was used in order to calculate the parameter if it existed within a bounded range; else a standard value was used for it.

4. The classification algorithms were presented with data from the same muscle location(Biceps Brachii) and signal type(Real or Simulated) but collected from different sources. In this case, the first source was the data acquired in the respective papers and its corresponding performance metrics. The second source was the data collected from same muscle location but the repository of Nikolic M.[41].

5. The classification algorithms was then presented with data of different amount in order to observe the change pf performance with increase in training data. The data was used from same repository(Nikolic M.). However, due to the insufficiency of open data for EMG signals of ALS and healthy subjects the highest size of dataset was limited to only 177 for real EMG signal and 100 from simulated EMG signal.

6. The Classification algorithms were finally presented with data acquired from the repository of Hamilton-Wright[48] which simulated EMG signals of Biceps Brachii of Neuropathic and Healthy subjects. The performance of this dataset was compared with the performance of real EMG signal obtained from clinical setup but from the same muscle location.

7. The performance of each classification technique was then measured using performance charts, Receiver Operating Characteristic(ROC) curve and Area Under the Curve(AUC) generated during classification.

8. All the algorithms were reconstructed according to the procedure described in respective papers.

3.2.5 Parameter Measurement

Time domain Parameters

A 6s long time-series EMG signal was used for extracting time domain features from the signal. The signal was lowpass filtered at 8KHz and high-pass filtered at 60Hz respectively. Noise was removed from the signal using DWT with a daubechies-4 wavelet and soft thresholding. 5 levels of decomposition was used for DWT.

Time-Frequency domain Parameters

Power Spectrum(PS), Fourier Coefficients and Power Spectral density(PSD) of the time-series signal was used for extracting features from the frequency and time-frequency domain of the signal. A 10ms long window was used in order to estimate AMD, VCF and VMF from spectrogram of the signal.

EMG Decomposition Parameters

EMG Decomposition was performed according to the procedure mentioned in [36]. The decomposition was performed for 8 motor units. This parameter was chosen since it has been mentioned in [36] that concentric needle electrodes can record the electrical activity of 5 to 10 motor units simultaneously. The minimum peak rise threshold for MUAP waveform was selected as 100 micro-volts as mentioned in [36]. The minimum threshold for potential rise duration was selected to be 0.5ms. This parameter was chosen in order to remove artifacts with high amplitude[36]. The MUAP waveform length was chosen to be 10ms.

3.3 System Architecture for "Multi-Domain Feature Extraction from EMG Signals and Evaluation for the Classification of Amyotrophic Lateral Sclerosis"

Electromyography(EMG) signals provide significant information for the diagnosis of neuromuscular disorders like Amyotrophic Lateral Sclerosis(ALS). Due to the stochastic nature of EMG signals different preprocessing and feature extraction techniques need to be applied in order to extract useful information from the raw noisy signals. Time-Frequency analysis and EMG Decomposition are two of the widely implemented techniques for feature extraction from EMG signals. However, due to extrinsic and intrinsic artifacts any one feature extraction technique alone does not provide enough information in order to show a consistent performance of classification across a variety of dataset. This paper proposes a method for classification of

ALS based on evaluation of multiple features extracted from three domains of EMG signal: time domain representation, frequency domain representation and Muscle Unit Action Potential(MUAP) waveform acquired via EMG decomposition of the signal. 43 features were evaluated using feature selection techniques like chi-squared test and recursive feature elimination.

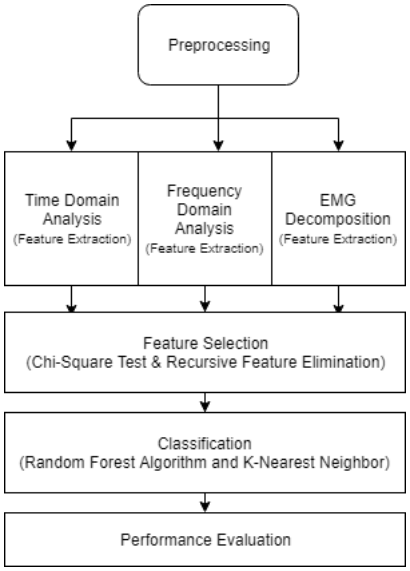

Figure 3.6: Block Diagram for classification using Multi-Domain Feature Extraction

Figure 3.6 shows a block diagram of the system architecture developed for the classification of ALS using feature extraction and selection from multiple domains.

3.3.1 Dataset

The dataset was collected from the repository of Nikolic M. which consisted of a normal control group and a group of patients with ALS.

3.3.2 Feature Extraction and Preprocessing

Signals in the dataset were first cropped from a total duration of eleven seconds to six seconds. Then DWT was applied in order to remove noise from the EMG signals. DWT applied with a soft thresholding is useful for removing artifacts from EMG signals[72], [75]. The denoised signal was later lowpass filtered at 8KHz using a second order butterworth filter in order to remove high frequency noise from the signal. The filtered signals were then passed on to the next stage for feature extraction.
Table 3.2 contains the features which were extracted via each of the aforementioned three feature extraction techniques. Full name along with details for the acronym of each feature is provided later in Chapter 4.

Time domain Features

Features extracted from time-series representation of EMG signals provide important information regarding probable abnormalities in the signal. Features like Mean Absolute Value, Cepstrum coefficient, entropy, Average Amplitude Change, etc. are

Time Domain	Time-Frequency Domain	EMG Decomposition
MAC	MDF	A.Dur
IMAC	MNF	V.Dur
RMS	TOP	A.Amp
STD	VCF	V.Amp
VAR	AMD	A.Ris
ENT	VMF	V.Ris
ZCR	ASA	A.Pha
ZLA	MNP	V.Pha
SKW	-	A.Are
KUR	-	V.Are
MAV	-	Min.Amp
WVL	-	Max.Amp
WLA	-	S.Amp
SSC	-	V.Amp
LDE	-	V.Min
MMAV	-	SCR
SMAD	-	ISI
-	-	FF

Table 3.2: Candidate Features extracted via Time-Series Analysis, Time-Frequency Analysis and EMG Decomposition

some of the useful time domain features that provide relevant information regarding electrical activity within the skeletal muscles of the body[97], [106]. 17 time domain features were extracted from each time-series signal and later passed on for feature selection.

Time-Frequency Features

Time-Frequency analysis is a signal processing technique which studies a signal both in time and frequency domain simultaneously using different time-frequency representations like power spectrum(PS), power spectral density(PSD), Short-Time Fourier Transform(STFT), etc[81]. This technique is quite better than frequency domain analysis because it retains some temporal information even after converting a time-series signal to its corresponding frequency-domain representation. We used PS and PSD of each EMG signal in order to extract 8 features provided in the table above.

EMG Decomposition Features

EMG signals recorded using needle electrode comprises of MUAP or spike trains generated from multiple motor units(MU) within the region of interest. Therefore, the signal contains rich information about the discharge pattern and organization of different MUs[49]. In order to extract information regarding different Mus, the signal needs to undergo a process known as EMG Decomposition which decomposes a signal into its probable constituent MUs by clustering groups of similar MUAP waveform based on their morphology[36]. Features like inter-spike-interval, rise time, duration, etc. can then be extracted from firing patterns and MUAP waveform obtained for

different MUs in order to perform classification. The EMG decomposition procedure which is implemented in this paper follows the algorithm presented by Christodoulou and Pattichis[36]. This algorithm was chosen due to its high decomposition accuracy and simplicity. Self-Organizing Feature Map(SOFM) was used for identification of similar MUAP waveform as described in [36]. 18 features were extracted from the morphology and firing pattern of MUAP waveform in order to perform classification.

3.3.3 Feature Selection

While extracting high level features from a low level data representation, it is likely that the data may contain irrelevant or redundant features that affect the analysis negatively. Even though feature extraction in pruning the data and identifying candidate variables, in many cases the size and dimensionality of scientific data make it difficult to use available domain information to identify features that discriminate between the classes of interest[46]. Feature Selection is a process of selecting some subset of a learning algorithm's input variables upon which it should focus attention, while ignoring the rest, also known as Dimensionality Reduction. There are different kinds of feature selection algorithms like Principal Component Analysis(PCA), Chi-Squared Test(CST), Recursive Feature Elimination(RFE), Ridge-Regression(RR), etc. The following two feature selection techniques were implemented in order to select the best subset of features from the set of candidate features extracted from each of the three domains:

Chi-Squared Test

Chi-Squared Test(CST) or Pearson's chi-squared test is a statistical test applied to sets of categorical data to evaluate how likely it is that any observed difference between the sets arose by chance. It is suitable for unpaired data from large samples[1], [79]. The equation for calculating CST is as follows:

$$\chi^2 = \frac{1}{d}\sum_{k=1}^{n}\frac{(O_k - E_k)^2}{E_k} \tag{3.1}$$

Here, 'O' is the number of observed count and 'E' is the number of expected counts and 'χ^2' is the measure of correlation between a feature and its corresponding class[33], [45], [94].
15 best features were selected in this paper using CST. The feature selection was performed using 5-fold Cross Validation and the best 15 features throughout all steps of the cross validation were selected as final feature set which was later used for classification.

Recursive Feature Elimination

Recursive Feature Elimination(RFE) is a feature selection technique that fits a model and rank features recursively according to some measure of importance. It uses the model accuracy to identify which attributes (and combination of attributes) contribute the most to predicting the target attribute. The pseudo-code for RFE is as follows: $T \leftarrow \{X_1, X_2, ..., X_n\}$ T is the training set consisting of n samples $F \leftarrow \{f_1, f_2, ..., f_p\}$ F is the set of p features $M(T, F)$ Ranking Method

$i = 1$ to p Rank set F using M(T,F) $f^* \leftarrow$ last ranked feature in F $R(p+1-i) \leftarrow f^*$ $F \leftarrow F - f^*$ The aforementioned pseudo code for RFE has been collected from [52]. 15 best features were selected in the paper using RFE. The feature selection was performed using 5-fold cross Validation and the best 15 features throughout all steps of the cross validation were selected as final feature set which was later used for classification.
Each of the 15 best features selected using the aforementioned techniques were passed on to the classifier individually for classification.

3.3.4 Classification and Performance Evaluation

In order to classify healthy subjects and subjects affected with ALS 15 best features were extracted from each signal in the dataset according to the procedure mentioned above. K-Nearest Neighbor(KNN) and Random Forest Algorithm(RFA) are the two supervised machine learning algorithm used in this experiment for classification.
RFA is an ensemble learning algorithm introduced by Ho which builds multiple trees in randomly selected sub-spaces which enables them to generalize their classification in complementary ways, and their combined classification can be monotonically improved of the feature space[24]. It implements stochastic discrimination approach to classification proposed by Eugene Kleinberg[17].
K-Nearest Neighbor Algorithm is a type of instance-based learning or non-generalizing learning: it does not attempt to construct a general internal model, but simply stores instances of the training data. Classification is computed from a simple majority vote of the nearest neighbors of each point: a query point is assigned the data class which has the most representatives within the nearest neighbors of the point. It is a non-parametric method used for classification and regression[19].
The hyper-parameters used in KNN and RFA are the number of neighbors and the number of decision trees respectively. The optimal hyper-parameter for each classifier was selected using Particle Swarm Optimization(PSO). It is a population-based search technique that utilizes the concept of social sharing of information[90]. The PSO algorithm which is implemented in this paper follows the procedure described in [90]. The dataset consisted of two output classes: normal subjects and ALS affected subjects. The performance of the classifiers are discussed in Chapter 4.

3.3.5 Parameter Measurement

Time domain Features

The following features were extracted from the time-series signal for classification:

1. **Maximum Autocorrelation(MAC):** Autocorrelation is the correlation of a signal with a delayed copy of itself as a function of delay. The maximum value of autocorrelation was selected as a feature as it provides information about periodic nature of the signal.

2. **Index of Maximum Autocorrelation(IMAC):** The index of maximum value of autocorrelation was selected as a feature as it provides information about the point where a signal is correlated with itself the most. It is usually at the 0^{th} index of autocorrelation.

3. **Root Mean Square(RMS):** The equation for RMS is as follows[81]:

$$RMS = \sqrt{\frac{1}{N}\sum_{k=1}^{N}(x_k)^2} \tag{3.2}$$

Here 'x' is a signal of length 'N'.

4. **Standard Deviation(STD):** The equation for STD is as follows[97]:

$$STD, \sigma = \sqrt{\frac{1}{N-1}\sum_{k=1}^{N}(x_k - \mu)^2} \tag{3.3}$$

Here 'x' is a signal of length 'N' and μ is mean of the signal.

5. **Variance(VAR):** The equation for VAR is as follows[97]:

$$VAR = (\sigma)^2 \tag{3.4}$$

6. **Zero Crossings(ZCR):** The equation for ZCR is as follows[81]:

$$ZCR = \sum_{n=1}^{N-1}[sgn(x_n * x_{n-1} \cap |x_n - x_{n+1}| \geq threshold] \tag{3.5}$$

Here 'x' is a signal of length 'N' and $sgn(x_i)$ is the sign of the value of 'x' at index 'i'.

7. **Zero Lag of Autocorrelation(ZLA):** The value at 0^{th} index or time of autocorrelation of the signal is ZLA of the signal.

8. **Skewness(SKW):** The equation for SKW is as follows[97]:

$$SKW = \frac{\frac{1}{N}\sum_{n=1}^{N}(x_n - \mu)^2}{\sigma^3} \tag{3.6}$$

9. **Kurtosis(KUR):** The equation for KUR is as follows[97]:

$$KUR = \frac{\frac{1}{N}\sum_{n=1}^{N}(x_n - \mu)^4}{\sigma^4} \tag{3.7}$$

10. **Mean Absolute Value(MAV):** The equation for MAV is as follows[65]:

$$MAV = \frac{1}{N}\sum_{k=1}^{N}|X_k| \tag{3.8}$$

Here, 'X' is a signal of length 'N'.

11. **Waveform length(WVL):** The equation for WVL is as follows[97]:

$$WVL = \sum_{n=1}^{N-1}|X_{n+1} - X_n| \tag{3.9}$$

12. **Willson Amplitude(WLA):** The equation for WLA is provided in [97].
13. **Slope Sign Change(SSC):** The equation for SSC is provided in [97].
14. **Log Detector(LDE):** The equation for LDE is provided in [65].
15. **Modified MAV(MMAV):** The equation for MMAV is provided in [81].
16. **Slope of Mean Absolute Deviation(SMAD):** The equation for Mean Absolute Deviation(MAD) is provided in [97]. SMAD is the average difference between MAD calculated using a 10ms window on the time-series signal.

Time-Frequency Features

The following features were extracted from frequency domain(power spectrum and Fourier coefficients) and time-frequency domain (power-spectral-density and power spectrum) of the signal:

1. **Mean Frequency(MNF):** The equation for MNF is provided in [81].
2. **Median Frequency(MDF):** The equation for MDF is provided in [81].
3. **Median Power(MNP):** The equation for MNP is provided in [81].
4. **Total Power(TOP):** The equation for TOP is provided in [81].
5. **Variance of Central Frequency(VCF):** The equation for VCF is provided in [81].
6. **Average Maximum Density(AMD):** Power-Spectral-Density(PSD) is the spectral energy distribution per unit time of the frequency components of a signal. AMD is the average of the maximum power-spectral-density(PSD) calculated with a 10ms window and no overlap.
7. **Variance of Maximum Frequency(VMF):** VMF is the variance of the maximum densities calculated over each 10ms window of the PSD.
8. **Average Spectral Amplitude(ASA):** ASA is the average of all peak amplitudes(above mean amplitude) of the Fourier Coefficients of a signal.

EMG Decomposition Features

The definitions for MUAP duration, amplitude, rectified area, rise time, phases, beginning point and ending point are provided in [36]. The following features were extracted from the average MUAP waveform and firing pattern calculated for each motor unit(MU) using EMG Decomposition:

1. **Average MUAP Duration(A.Dur):** A.Dur is the average of all the duration calculated for average MUAP waveform of each MU of a decomposed EMG Signal.
2. **Variance of MUAP Duration(V.Dur):** V.Dur is the variance of all the duration calculated for average MUAP waveform of each MU of a decomposed EMG Signal.

3. **Average MUAP Amplitude(A.Amp):** A.Amp is the average of all the amplitudes calculated for average MUAP waveform of each MU of a decomposed EMG Signal.

4. **Variance of MUAP Amplitude(V.Amp):** V.Amp is the variance of all the amplitudes calculated for average MUAP waveform of each MU of a decomposed EMG Signal.

5. **Average MUAP Rise Time(A.Ris):** A.Ris is the average of all the rise time calculated for average MUAP waveform of each MU of a decomposed EMG Signal.

6. **Variance of MUAP Rise Time(V.Ris):** V.Ris is the variance of all the rise time calculated for average MUAP waveform of each MU of a decomposed EMG Signal.

7. **Average MUAP Phase(A.Pha):** A.Pha is the average of all the phases calculated for average MUAP waveform of each MU of a decomposed EMG Signal.

8. **Variance of MUAP Phase(V.Pha):** V.Pha is the variance of all the phases calculated for average MUAP waveform of each MU of a decomposed EMG Signal.

9. **Average MUAP Rectified Area(A.Are):** A.Are is the average of all the rectified areas calculated for average MUAP waveform of each MU of a decomposed EMG Signal.

10. **Variance of MUAP Rectified Area(V.Are):** V.Are is the variance of all the rectified areas calculated for average MUAP waveform of each MU of a decomposed EMG Signal.

11. **Minimum Amplitude(Min.Amp.):** Min.Amp. is the minimum amplitude obtained from the average waveform calculated from the MUAP waveform of each MU.

12. **Maximum Amplitude(Max.Amp.):** Max.Amp. is the maximum amplitude obtained from the average waveform calculated from the MUAP waveform of each MU.

13. **Slope between Minimum and Maximum Amplitude(S.Amp.):** S.Amp. is the difference between Min.Amp. and Max.Amp.

14. **Variance between Maximum and Minimum Amplitude(V.Amp.):** V.Amp. is the variance of MUAP waveform between Min.Amp. and Max.Amp.

15. **Variance between Beginning Point and Minimum Amplitude(V.Min.):** V.Min. is the variance of the MUAP waveform between beginning point and minimum amplitude of the average MUAP waveform obtained from the waveform of each MU.

16. **Spike-Count-Rate(SCR):** SCR is obtained by counting the number of action potentials or spikes that appear during a trial in a spike train and dividing the result by duration of the trial. The spike train was obtained from firing sequence of the different MU via EMG Decomposition.

17. **Mean Inter-Spike-Interval(ISI):** ISI is the average time duration between two spikes in spike train of a decomposed EMG signal.

18. **Fano Factor(FF):** Fano factor is one of the most widely used measures of variability of spike trains. Its standard estimator is the ratio of sample variance to sample mean of spike counts(SCR) observed in a time window[93]. A 10ms window was used to calculate FF from the spike train of a decomposed EMG signal.

Chapter 4

Experimental Results

This chapter contains the experimental results obtained from the classification techniques used in Section 3.2 and 3.3. Section 4.1 contains the results obtained from Section 3.2 followed by a detailed analysis of the whole classification procedure along with its results. Similarly, Section 4.2 contains a detailed analysis report for the classification algorithm proposed in Section 3.3.

4.1 Experimental results for "Review of different EMG feature extraction and classification techniques and their limitations during the development of a general purpose application software"

This section of the paper presents a brief description of procedures followed in order to reconstruct the algorithms being analyzed, result obtained in the respective papers and the results obtained by applying the same classification techniques on different dataset.

4.1.1 Classification using Time-Frequency Analysis

EMG signal being stochastic in nature contains a lot of noise information due to various factors[103]. It is necessary to extract the useful information out of raw EMG data which requires rather a complex set of statistical analysis technique. One of them is to extract information from time domain as well as from frequency domain of a preprocessed signal[77]. Some of the common statistical features extracted from the time domain of a signal are MAV, Maximum Scatter Difference, Root Mean Square, SD or Variance, Approximate Entropy, Autoregressive coefficients, Cepstral coefficients, Histogram, Higuchi's Fractal Dimension, Kurtosis, etc. Moreover, the statistical features extracted from the frequency domain of the signal requires some kind of transformation of the signal to its corresponding frequency domain using techniques like Fourier Transform[10] introduced by Joseph Fourier, DWT[39] introduced by Alfred Haar, etc. The common frequency domain features are MNF, Median Frequency, Peak Frequency, Spectral Moment, Spectral Amplitude, etc.

These extracted features are used as input to various classification algorithms like KNN, SVM, etc.

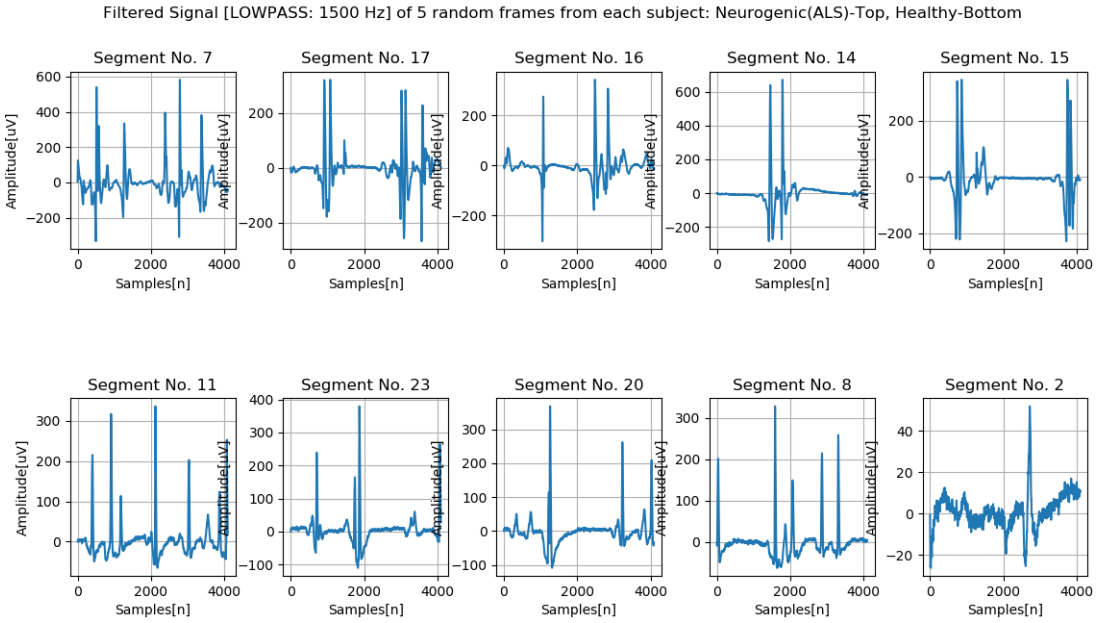

Figure 4.1: Filtered Segments/Frames

Figure 4.1 consists five random segments from the filtered EMG signal of a Neuropathic and a Healthy Subject. The signal was lowpass filtered at 1500Hz. Figure 4.2 shows the corresponding output of Magnitude spectrum for each filtered segment using FFT for a neuropathic(Top) and a healthy subject(bottom). Figure 4.3 shows the corresponding output of Autocorrelation for each filtered segment for a neuropathic(Top) and a healthy subject(bottom).

SL No.	**Type**	**Maximum**	**Minimum**	**Average**
0	Neuropathic	0.76155e5	0.975e3	0.13761e5
1	Healthy	0.105759e6	0.751e3	0.10104e5
2	Neuropathic	0.41662e5	0.54e2	0.4866e4
3	Healthy	0.16210e5	0.296e3	0.2812e4
4	Healthy	0.105762e6	0.851e3	0.7392e4
5	Neuropathic	0.33692e5	0.198e3	0.4863e4

Table 4.1: Spectral Amplitude of 3 Neuropathic(ALS) and 3 Healthy Patients using Magnitude Spectrum of FFT

Table 4.1 shows the Maximum, Minimum and Average ASP obtained from the filtered EMG signal of 6 random subjects using magnitude spectrum of FFT.
Figure 4.4 shows performance graph for classification of EMG signal using ASP of each signal frame as a feature. X-axis of the graph represents the number of dataset used for training and evaluating each new instance of a classifier and Y-axis shows the performance unit in percentage. The blue line shows change in test accuracy of the classifier as the number of training data was increased. Similarly orange line shows the change of validation or evaluation accuracy using unknown

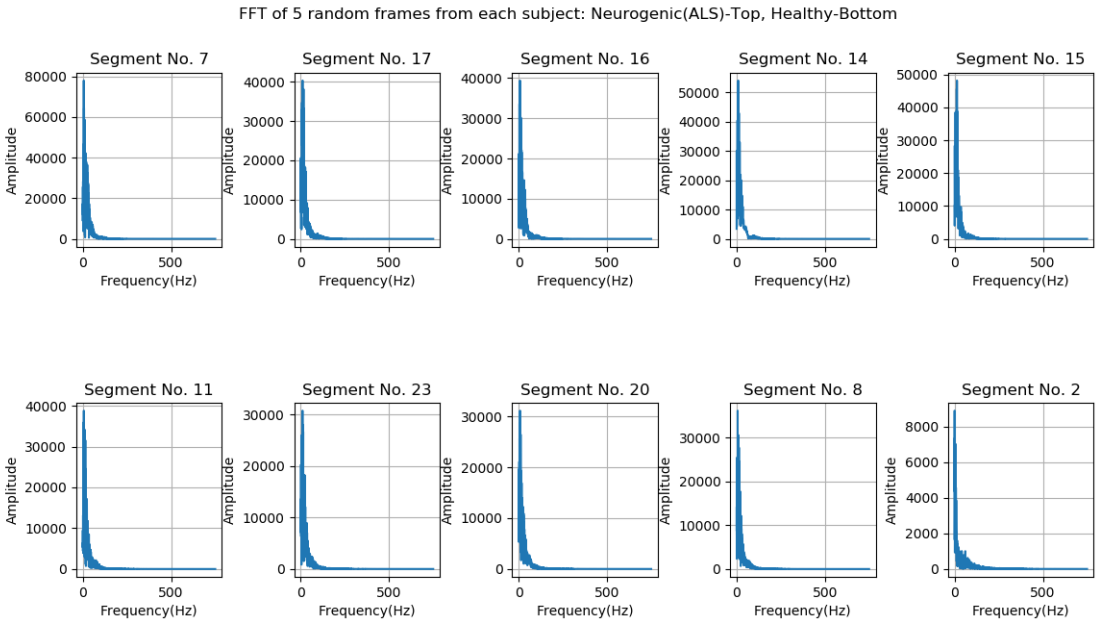

Figure 4.2: Magnitude Spectrum of Fast Fourier Transform

data, purple line shows the number of Nearest Neighbors used for classification of the input dataset, red line shows the change in Specificity of the classifier and green line shows change in Sensitivity of the classifier. KNN has been used for performing classification where the number of neighbors are optimized with Particle Swarm Optimization. The range of performance is from 0 to 100 percent and the range of input training dataset is from 0 to 177. It can be observed from the figure that even though training or test accuracy of the classifier fluctuates within 60%-70% with an increase of input dataset, the validation or evaluation accuracy increases slightly from 65% to 80%. The variation of training and test accuracy generally occurs due to over-fitting of training data when the model adapts to noise and detail such that it negatively impacts the performance of the model on new data. On the other hand, reduction of accuracy of test data with increase in number of classification input data shows the degradation in performance of the classifier while handling large dataset.

Table 4.2 shows the average performance metrics(Average Accuracy-AAC, Average Specificity-ASP, Average Sensitivity-ASE) for classification with different input features(for real and simulated EMG signal) over varying amount of input dataset. The values range from 0 to 100%. Here 'Feature' column represents the input feature used for classification. The input features are Average Amplitude of Spectral Peak(ASA), Mean Frequency(MNF), Zero Lag of Autocorrelation(AZL) and Zero Crossing rate(ZCR). Here, 'PSO' indicates that Particle Swarm Optimization(PSO) was used for choosing best fitted nearest neighbors(else 1 Nearest Neighbor was used). The input dataset contained a maximum of 177 data of Neuropathic and Healthy patients and 99 data of Simulated EMG signals of Neuropathic and Healthy patients. Ten fold cross validation is used for testing and validating the classifier.

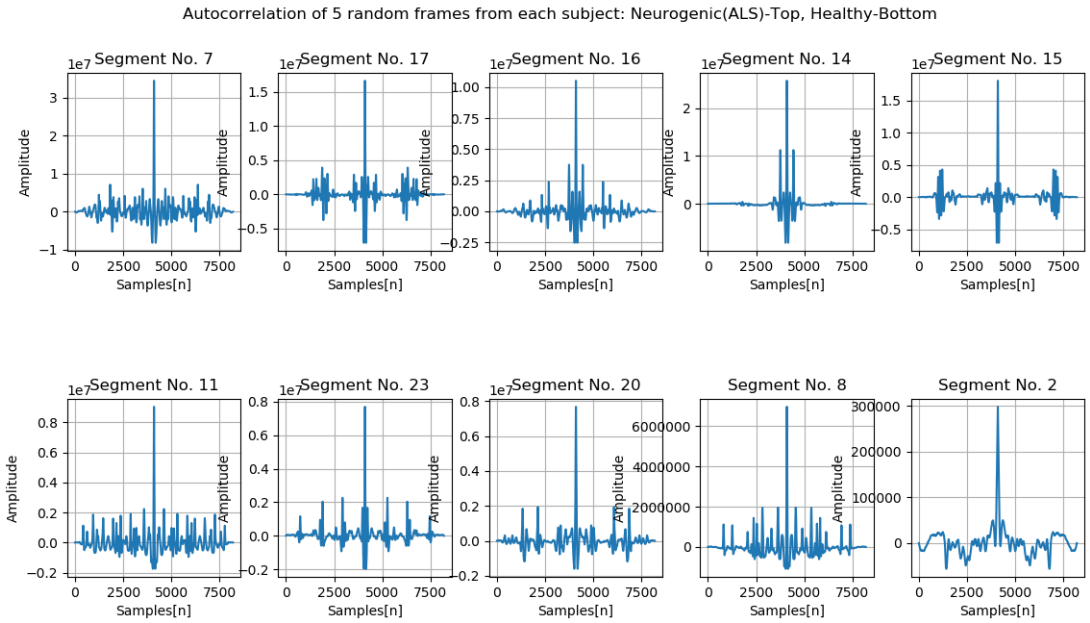

Figure 4.3: Autocorrelation

4.1.2 Classification using Discrete Wavelet Transform and Particle Swarm organization

One of the approaches for classification of EMG Signals is DWT. It is a wavelet transformation technique initially introduced by Haar that divides a signal into sub-bands of different frequency resolution. These sub-bands or basic functions are known as wavelets[90]. Like Fourier Transform, it helps to analyze a signal in its frequency domain. However, one of its advantages over Fourier Transform is that it has a changeable window dimension i.e. for lower frequencies it is wide and for higher frequencies it is narrow and hence provides a maximum frequency resolution at all intervals[61]. One of the constraints of DWT is that a predefined mother wavelet needs to be chosen for decomposition. There are different kinds of mother wavelets and db4 wavelet is chosen in the paper for decomposition. A five level decomposition is used which produces five details and one approximation coefficient. These coefficients are a list of values at different frequency resolution of the signal. Features are extracted from these six set of values for each EMG signal. This feature extraction technique is applied in order to reduce the dimensionality without losing much important information.

Table 4.3 shows the average performance metrics(Average Accuracy-AAC, Average Specificity-ASP, Average Sensitivity-ASE) for classification with different input features over varying amount of input dataset. The 'Classifier' Column represents the classifier used for performing classification. SVM with RBF Kernel(SVM-RBF), SVM with Polynomial Kernel(SVM-POLY) and KNN classifiers are used here. Fewer number of EMG signal data were used in this experiment(177 for real signals and 99 for simulated signals) compared to 1200 as mentioned in the paper in order to observe the performance of the algorithm for reduced number of data. DWT Features were used for classification.

4.5 and 4.6 shows the Performance Graph for classification of real and simulated

SL No.	Feature	AAC(%)	ASP(%)	ASE(%)
1	ASA(Test, Real)	68.22	18.66	57.10
2	MNF(Test, Real)	43.39	66.90	52.77
3	AZL(Test, Real)	74.63	25.43	73.19
4	ZCR(Test, Real)	71.69	37.89	84.62
5	ASA(Validation, Real)	68.22	63.11	69.92
6	MNF(Validation, Real)	48.01	52.41	46.50
7	AZL(Validation, Real)	58.43	69.05	54.56
8	ZCR(Validation, Real)	49.56	32.12	69.99
9	ASA(Test, Simulated)	72.40	100	92
10	MNF(Test, Simulated)	89.22	90.47	98.21
11	AZL(Test, Simulated)	78.43	100	100
12	ZCR(Test, Simulated)	66.73	90.47	87.44
13	ASA(Validation, Simulated)	69.58	0	90.07
14	MNF(Validation, Simulated)	66.43	4.76	92.85
15	AZL(Validation, Simulated)	77.80	0	93.57
16	ZCR(Validation, Simulated)	85.91	0	96.82

Table 4.2: Average performance table for classification of Neuropathic and Healthy patients based on different test and validation data(TFA Features).

SL No.	Classifier	AAC(%)	ASP(%)	ASE(%)
1	SVM-RBF(Test, Real)	71.46	32.53	75.08
2	SVM-POLY(Test, Real)	52.97	100	100
3	KNN(Test, Real)	76.87	35.51	93.19
4	SVM-RBF(Validation, Real)	74.32	55.91	80.59
5	SVM-POLY(Validation, Real)	57.32	28.57	71.42
6	KNN(Validation, Real)	77.11	59.52	85.71
7	SVM-RBF(Test, Simulated)	90.34	100	96.42
8	SVM-POLY(Test, Simulated)	72.54	97.14	100
9	KNN(Test, Simulated)	68.004	82.14	92.85
10	SVM-RBF(Validation, Simulated)	68.35	0.0	100
11	SVM-POLY(Validation, Simulated)	70.18	0.0	100.0
12	KNN(Validation, Simulated)	72.40	0.0	93.65

Table 4.3: Average performance table for classification of Neuropathic and Healthy patients based on different test and validation data(DWT Features) using real and simulated EMG signal.

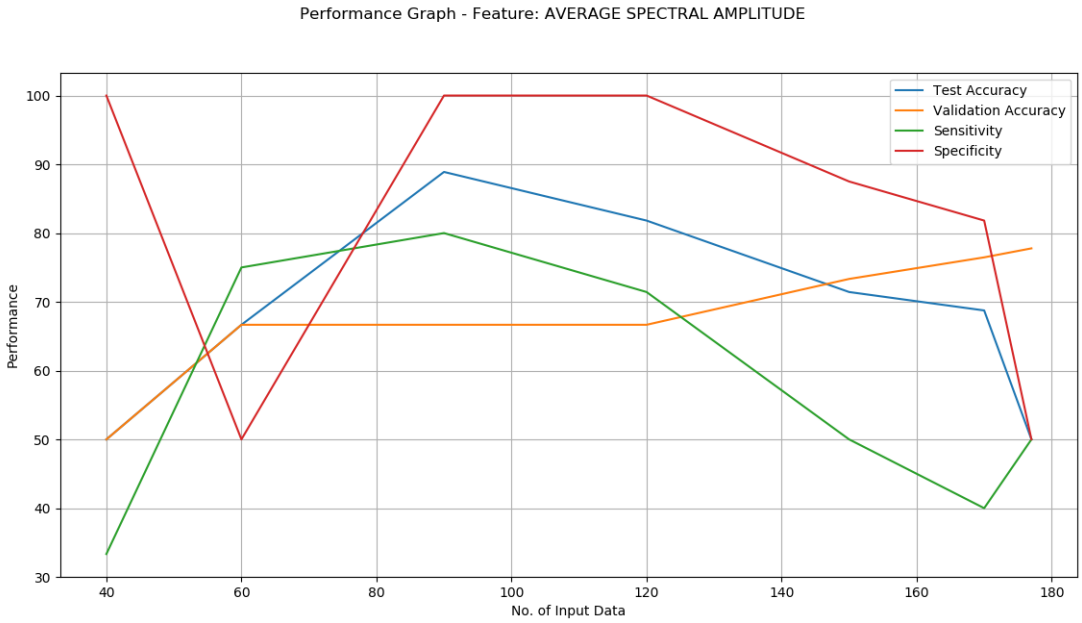

Figure 4.4: Performance graph for EMG Classification of Neuropathic(ALS) and Healthy subjects.

EMG signal using DWT, SVM with RBF kernel and PSO. The x-axis represents the number of input data provided to the classifier and y-axis represents the percentage of Test Accuracy, Validation Accuracy, Specificity and Sensitivity of the classifier. Figure 4.7 shows the ROC Curve and corresponding AUC for classification performance of KNN classifier. MAV, SD, AP and RMAV were extracted as features from the coefficient sub-bands of DWT.
The performance graph shows an average increase in performance for increasing dataset size than TFA. However, the average accuracy for smaller dataset size shows a decrease in classification accuracy which varies from 70%-80% for small dataset of 177 patient records compared to that of 1200 patient records[90] which shows an accuracy above 90%.

4.1.3 Classification using unsupervised pattern recognition and EMG Decomposition technique

One of the approaches for classification of EMG signal is decomposition of the signal into its constituent MUAP and firing pattern for estimated number of motor units(MU). These information(waveform of the MU and firing table) are then used to identify different abnormalities in the EMG signal. This approach mimics the traditional manual form of identification of diseases in clinical sectors where an expert usually identifies different probable MUAP waveform by observing the signal manually and template matching techniques. Algorithms proposed by Dorfman[13] and LeFever[9] are some of the examples of manual or operator dependent method of EMG decomposition. However, manual form of EMG decomposition is not only time consuming but generally provides only 1% or less information about few MUAP waveform[21]. However, such small percentage of information still provides a high accuracy for diagnosis. Automated EMG decomposition technique loosely mimics

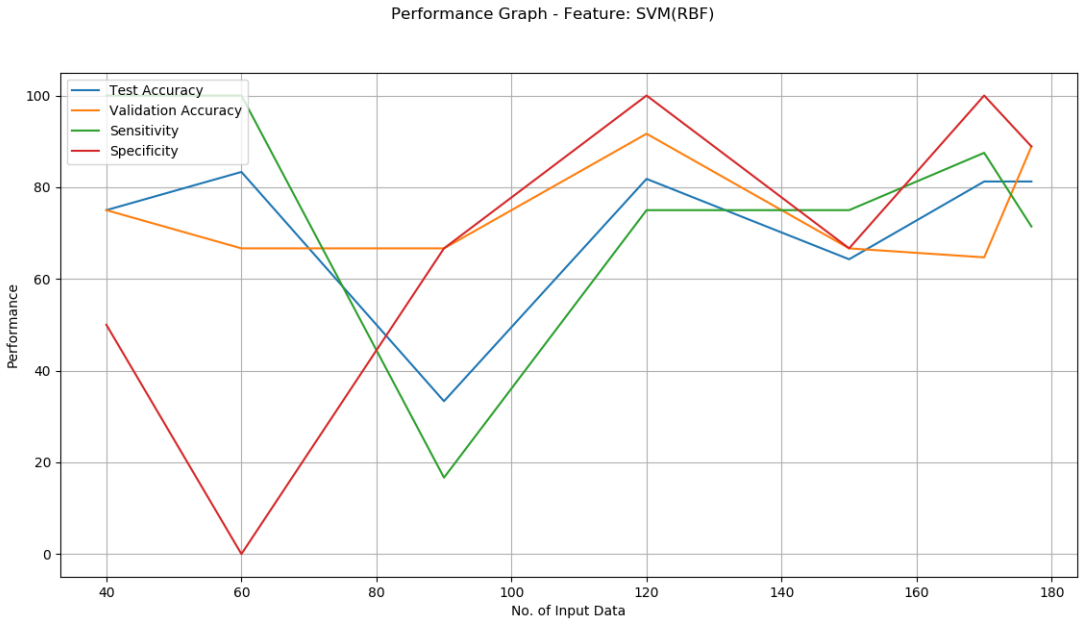

Figure 4.5: Performance Graph for Classification of real EMG Signal using Support Vector Machine with RBF Kernel for varying length of input dataset

the manual operator dependent procedures followed in order to decompose a signal into its constituent MUAP waveform and firing patterns and then perform diagnosis by extracting different features like phase, number of turns, etc. from these waveform. The algorithm proposed Christodoulos et. al.[36] uses SOFM which is a neural network model and LVQ2 method in order to decompose a raw EMG signal into their constituent(actual and superimposed) MUAP for a predefined number of MUs. The superimposed waveform are then decomposed using the template obtained from the actual MUAP waveform in order to construct a firing table for the motor units. Features like Turns, Number of Phases, Rectified Area, Duration and Amplitude are then calculated from average MUAP waveform of each motor unit in order to perform classification of the signal.

Table 4.4 shows the average performance metrics(Average Accuracy-AAC, Average Specificity-ASP, Average Sensitivity-ASE) for the classification of the decomposed EMG signals using different classifiers for an increasing number of dataset. The classification was performed using 177(7 neuropathic and 3 healthy subjects) real EMG signals obtained from biceps brachii and 99(15 Neuropathic and 5 Healthy subjects) simulated signals.

Figure 4.8, 4.9 and 4.10 shows the MUAP waveform decomposed from the EMG signal, the average MUAP waveform for each MU along with their extracted features and the firing table constructed using the EDA respectively. It was presented by Christodoulos et. al[36]. Figure 4.8 contains eight plots where each plot displays the different MUAP templates obtained from the signal which is expanded according to the technique mentioned in the paper. Figure 4.9 contains the average MUAP class calculated from the previous templates of each motor unit. The red, green and blue cross marks show the Beginning Extraction Point(BEP), End Extraction Point(EEP) and the waveform Peaks calculated in order to obtain the desired features from each motor unit. The title bar of each motor unit plot con-

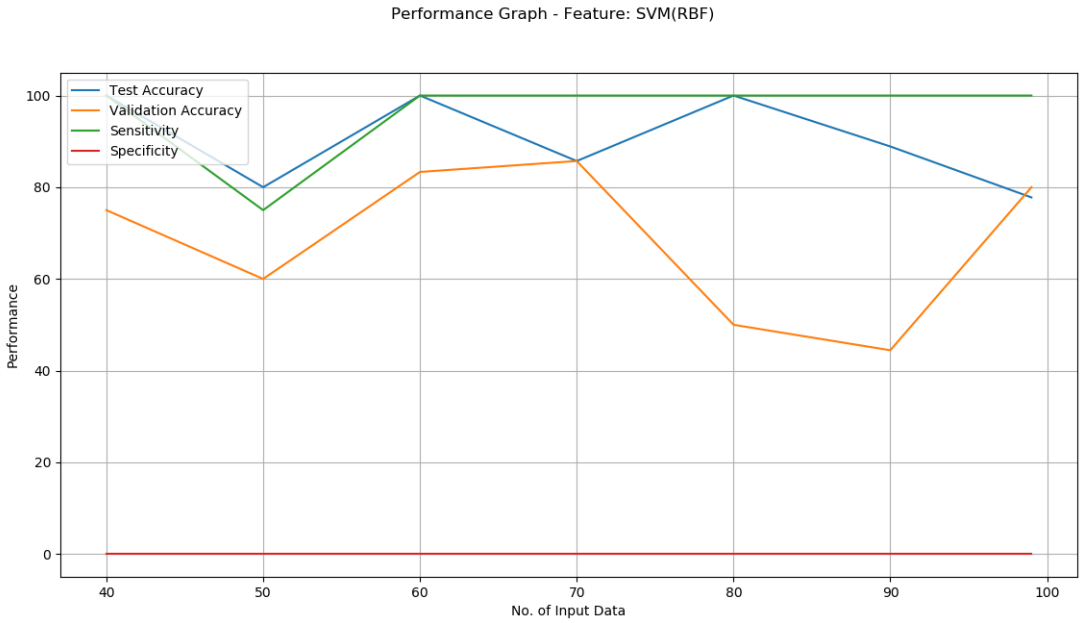

Figure 4.6: Performance Graph for Classification of simulated EMG Signal using Support Vector Machine with RBF Kernel for varying length of input dataset

tains the Peak to Peak Amplitude(A), Duration(D), Rectified Area(Ar), number of turns(T) and Phases(P) of the average MUAP waveform. Figure 4.10 shows the candidate MUAP waveform(Top) calculated using peak detection and thresholding technique before MUAP segmentation and classification. Red Waveform represents the candidate MUAP waveform detected and the green mark on top represents the peak from which the waveform was obtained. It also shows the firing table calculated(Middle) for each motor unit before the decomposition of superimposed MUAP and the bottom graph shows the final firing table obtained after the decomposition of the superimposed waveform using the template MUAP waveform obtained during segmentation. Each color of the waveform represents a different Motor Unit class. Figure 4.11 and 4.12 shows the Performance Graph for classification of real and simulated EMG signal using EDA, KNN and PSO. The x-axis represents the number of input data provided to the classifier and y-axis represents the percentage of Test Accuracy, Validation Accuracy, Specificity and Sensitivity of the classifier.

4.1.4 Discussion

Electromyography is a very useful technique for diagnosis of different neuromuscular disorders. Despite of the stochastic nature and inherent noises of EMG signals, it is very effective even with the small amount of information which is currently retrieved from it. However, the techniques used for detection of these disorders are very much dependent on operator. A software which is able to automate this process of signal classification can provide a fast and detailed insight about the significant differences between nerve activity of neuropathic, myopathic and healthy subjects.
But one of the important properties of a good software is reliability. It needs to be adaptive to different kind of dataset which means that the output accuracy should be similar even when the analysis is performed for different set of data like from dif-

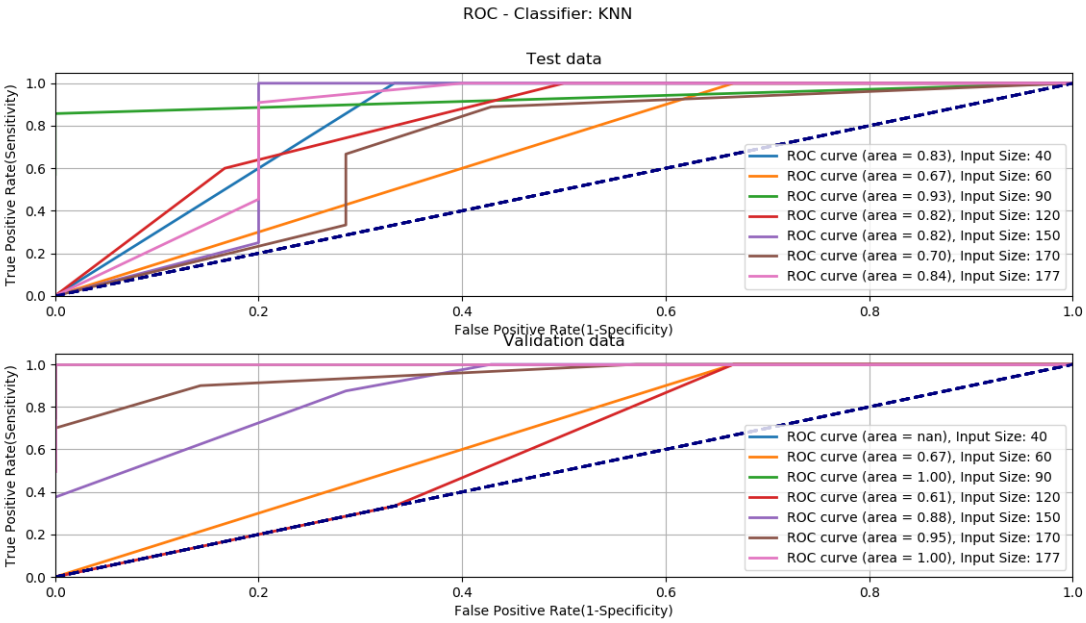

Figure 4.7: ROC Curve for K Nearest Neighbor using real signal

ferent sources, different signal type and even with different number of training data as it is impractical for every user to acquire a large training dataset before starting to perform analysis. Different kind of techniques(for needle electrode EMG) for classification of neuropathic, myopathic and healthy subjects can be generalized into three basic types: Classification of signal using TFA of the Signal, classification of the signal using DWT and classification using EDA. The previous section presented performance metrics of three classification algorithms each belonging to a different classification technique. This section discusses about the performance of these classification algorithms when it is provided with training data(EMG signals using Needle Electrode from biceps brachii) having three different properties:

1. The classification algorithms are reconstructed in a similar way as described in the respective papers but using training dataset acquired from open source data repositories. This factor was considered as the development of an artificial intelligence requires constant testing and often with readily available dataset. In this age of collaboration and networking, one of the best ways to acquire sufficient data is from open source data repositories.

2. The number of training data is either larger or smaller than the training dataset used in the classification algorithms.

3. The type of training data is changed from real clinical EMG signals to simulated experimental EMG signals.

Table 4.5 shows the performance chart of TFA, DWT and EDA for classification of EMG signals of healthy and neuropathic subjects. The performance metrics(accuracy, specificity and sensitivity) for each classification technique is divided into 3 main columns: Metrics obtained from real clinical EMG signal(Real), metrics obtained from simulated EMG signal(Sim) and the average performance metric for varying amount of training data(Avg). TFA and DWT classification techniques

SL No.	**Feature**	**AAC(%)**	**ASP(%)**	**ASE(%)**
1	SVM-RBF(Test, Real)	67.82	64.28	84.64
2	SVM-POLY(Test, Real)	46.98	85.71	100
3	KNN(Test, Real)	60.71	55.0	80.25
4	SVM-RBF(Validation, Real)	74.32	55.91	80.59
5	SVM-POLY(Validation, Real)	72.33	51.83	91.66
6	KNN(Validation, Real)	58.33	14.28	71.42
7	SVM-RBF(Test, Simulated)	66.31	100.0	100.0
8	SVM-POLY(Test, Simulated)	81.16	100.0	100.0
9	KNN(Test, Simulated)	80.69	100.0	91.51
10	SVM-RBF(Validation, Simulated)	68.34	0.0	100.0
11	SVM-POLY(Validation, Simulated)	86.12	0.0	100.0
12	KNN(Validation, Simulated)	75.65	0.0	91.50

Table 4.4: Average performance table for classification of Neuropathic and Healthy patients based on different test and validation data(EDA Features) using real and simulated EMG signal.

Metric (%)	TFA				DWT				EDA		
	Exp	Real	Sim	Avg	Exp	Real	Sim	Avg	Real	Sim	Avg
Accuracy	100(1)	68.2	69.6	68.4	95.2(a)	77.1	72.4	77.7	58.3(a)	75.6	68.1
	69.5(2)	48.1	66.4	42.5	96.7(b)	57.32	70.2	57.6	72.3(b)	86.1	86
	100(3)	58.4	77.8	71.1	97.4(c)	74.3	68.3	73.6	74.3(c)	68.3	60
	72.2(4)	49.6	85.9	73.8	-	-	-	-	-	-	-
Specificity	100(1)	63.11	0	81.4	92(a)	59.5	0	65.4	14.3(a)	0	45.7
	66.7(2)	52.4	4.8	32.8	93.5(b)	28.6	0	0	51.8(b)	0	0
	100(3)	69.1	0	75	95.3(c)	55.9	0	68.6	55.9(c)	0	42.8
	72.2(4)	32.1	0	63	-	-	-	-	-	-	-
Sensitivity	100(1)	69.9	90.1	57.4	96.7(a)	85.7	93.6	92.8	71.4(a)	91.5	81.4
	72.2(2)	46.5	92.8	54	99(b)	71.4	100	100	91.7(b)	100	100
	100(3)	54.6	93.6	73.6	99(c)	80.6	100	75	80.6(c)	100	90.7
	72.2(4)	70	96.8	84.7	-	-	-	-	-	-	-

Table 4.5: Comparison of performance of 3 different classification techniques(TFA, DWT and EDA) for different type of dataset containing EMG records of healthy and neuropathic subjects.

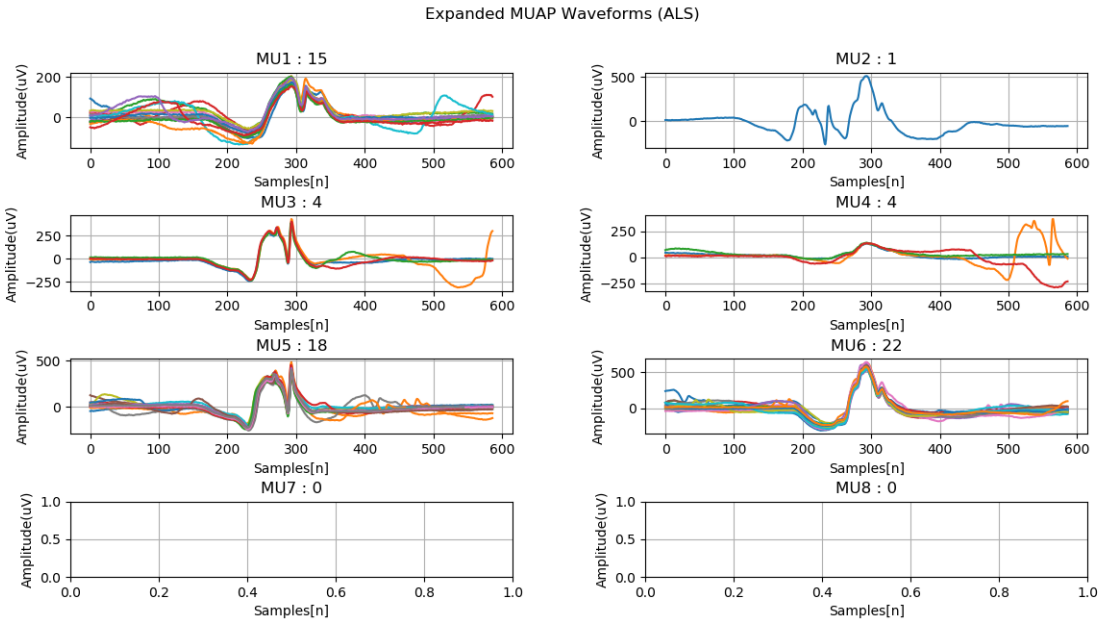

Figure 4.8: MUAP Waveform decomposed and extracted for each Motor Unit of a neuropathic patient affected by Amyotrophic Lateral Sclerosis.

contain one additional column(Exp) that shows the performance metrics obtained in the respective papers. The algorithm proposed by Christodoulos et. al[36] did not perform classification of EMG signals for neuromuscular disorder and hence it does not contain any column for 'Exp'. The performance metric obtained for each parameter(accuracy, specificity and sensitivity) in TFA are divided into four rows(1, 2, 3 and 4). Each row contains the performance metric for classification using four different features: Spectral Peak(1), Mean Frequency(2), Autocorrelation(3) and Zero Crossing rate(4). On the other hand, performance metric obtained for each parameter in DWT and EDA are divided into three rows(a, b and c). Each row contains the performance metric for classification using four different classifiers: KNN(a), SVM with polynomial kernel(b) and SVM with RBF kernel and optimized using PSO(c). The classification of TFA analysis was performed using KNN.
The following facts can be observed from the performance chart of different EMG classification techniques for detection of neuropathy (Amyotrophic Lateral Sclerosis) and healthy subjects:

1. The performance accuracy for each classification technique(TFA and DWT) varied considerably when the same technique was used on dataset obtained from different repositories(Nikolic M., Hamilton-Wright and dataset used in the original papers). This fact can be observed by comparing the performance metric of each of the four columns of TFA and DWT analysis technique.

2. The performance metric of each classification technique(TFA and DWT) showed a better performance when the classification was performed in respective papers of each analysis technique(Exp or experimental dataset column). The specificity and sensitivity for experimental dataset columns showed a better performance when classification was performed using DWT. On the other hand, even though optimal sensitivity was obtained, the algorithms provided

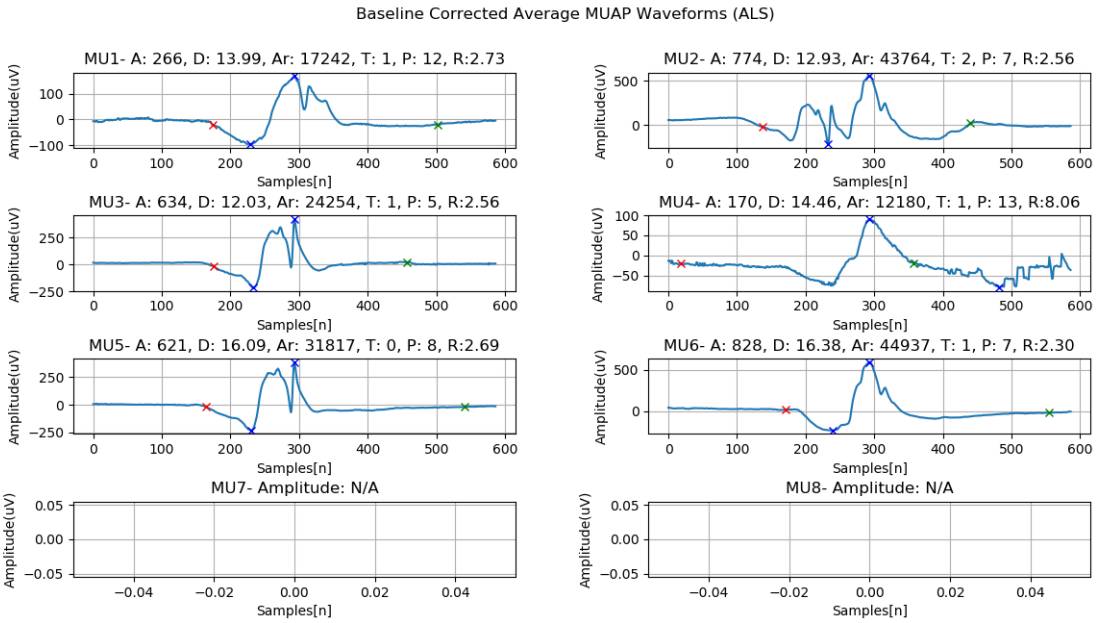

Figure 4.9: MUAP Waveform averaged and baseline rectified for each Motor Unit of a neuropathic patient affected by ALS.

poor specificity metrics when classification was performed with real and simulated signal.

3. DWT and EDA analysis techniques perform better when classification is performed using real EMG signals(Real column). However, the accuracy is below 90% when classification is performed using signals obtained from the repository of Nikolic M.

4. EDA and TFA analysis techniques show a better accuracy when classification is performed using simulated EMG signals(Sim column) obtained from the repository of Hamilton-Wright. However, all the three classification techniques show poor specificity for simulated signals but a high sensitivity for neuropathic patients.

5. The classification accuracy for each technique(TFA, DWT and EDA) varies within a range of 50% to 80% for clinical EMG signals(Real) obtained from biceps brachii using a needle electrode[41]. On the other hand, when the same classification technique was applied in the experiment(TFA) it displayed an accuracy of 100% for Spectral Peak(1) and Autocorrelation(3). This shows the variation of performance when classifying stochastic signals like EMG are obtained from different data acquisition setup or data repositories. It is one of the primary factors that acts as a barrier when trying to develop an adaptable software. Although this problem can be reduced by adjusting the parameters of described algorithms, this increases the need for operator assistance and reduces the prospect for automation. Similarly, the classification accuracy for DWT analysis reduced from range of ninety percent and above to around 70-80%. The reduction of specificity and sensitivity of the algorithms can also be observed in the same way.

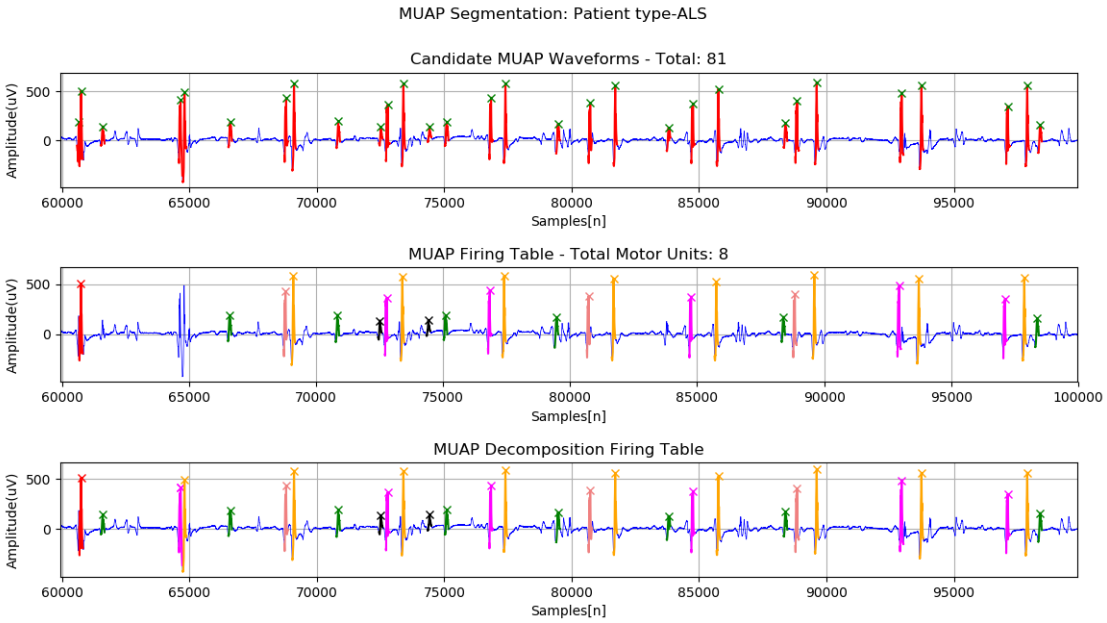

Figure 4.10: Firing Table constructed for each motor unit using decomposed MUAP waveform of a patient affected by Amyotrophic Lateral Sclerosis.

6. The classification of simulated EMG signals provides a better performance metric than classification of real clinical EMG signals. Both signal categories(real and simulated) represent the electrical activity of biceps brachii of healthy and neuropathic patients obtained using a needle electrode. One of the probable reason for this might be the representation of signal to noise ratio(SNR) for real and simulated EMG signals. The nature of noise among acquired data is likely to vary greatly when obtained from clinical setup than simulations. The variation of performance metric can be observed from the 'Real' and 'Sim' column of each(TFA, DWT and EDA) classification technique. The highest accuracy obtained for real signals is 77.1%(DWT) and for simulated signals is 86.1%(EDA).

7. EDA provides a better performance metric when classification is performed using different amount of input or training data(Avg column). The highest average accuracy for varying amount of input data is 86%. It was obtained when classification was performed for real signals using SVM and Polynomial Kernel(b). The chart shows that EDA provides a good metric for average sensitivity while providing a poor metric for average specificity. This shows the reduction of performance of the classification algorithms when provided with small amount of dataset. This causes a problem during development of the automated software as it requires a large training dataset which is not always available due to aforementioned causes.

8. The sensitivity of the classification techniques(TFA, DWT and EDA) is generally better than their specificity. This shows that the algorithms provide a good probability metric while identifying a neuropathic patient when the disease is present(sensitivity). On the other hand, it provides a poor probability metric for True Negative(specificity). One of the probable reason for the

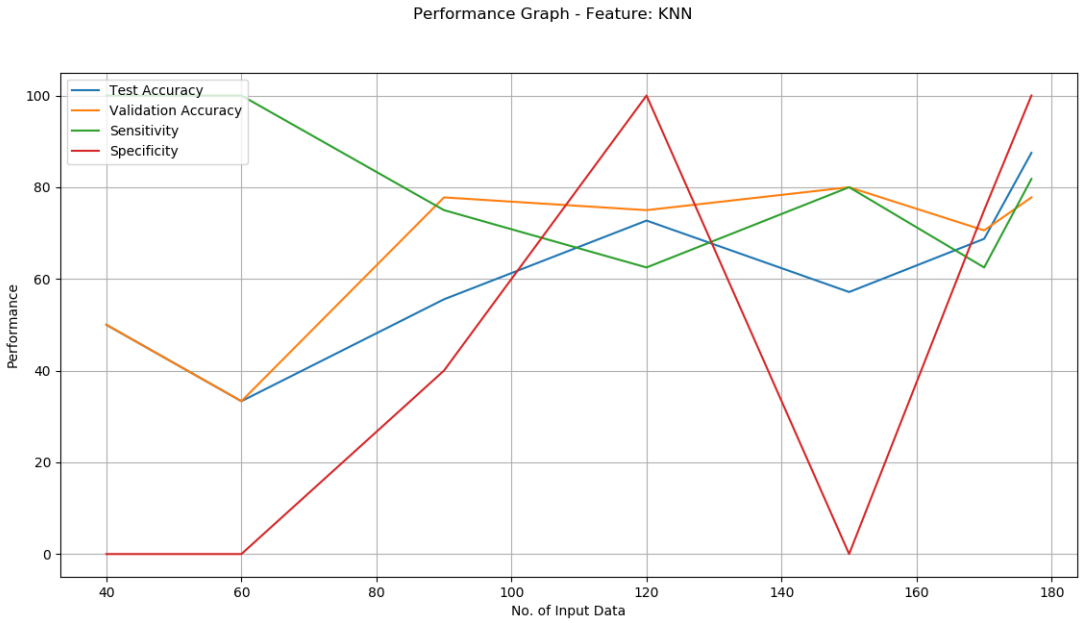

Figure 4.11: Performance Graph for Classification of real EMG Signal using K-Nearest neighbor for varying length of input dataset.

contrast between specificity and sensitivity might be the usage of unbalanced dataset which will generally be the case when trying to build an adaptable software. This factor can be controlled by balancing the dataset but it would depend more on trimming specific amount of data rather than adding it which would reduce the number of input or training data and again increase the need for manual preprocessing of data.

9. classification of EMG signals using different machine learning algorithms depend greatly on the nature and attributes of input or training data provided to the classifiers for training. This results in varying outcome when performing classification with a specific technique in different clinical or experimental setups. An automated software for general purpose usage needs to be robust and adaptable to different kinds of training data. This requires algorithm that can perform in the same way even when the aforementioned attributes of the training data are changed in order to minimize the need for human intervention.

4.2 Experimental results for "Multi-Domain Feature Extraction from EMG Signals and Evaluation for the Classification of Amyotrophic Lateral Sclerosis"

This paper proposes a multi-domain feature extraction method for the classification of EMG signals acquired from healthy subjects and subjects affected with ALS. The EMG signals are first resized, denoised using DWT and lowpass filtered at 8KHz

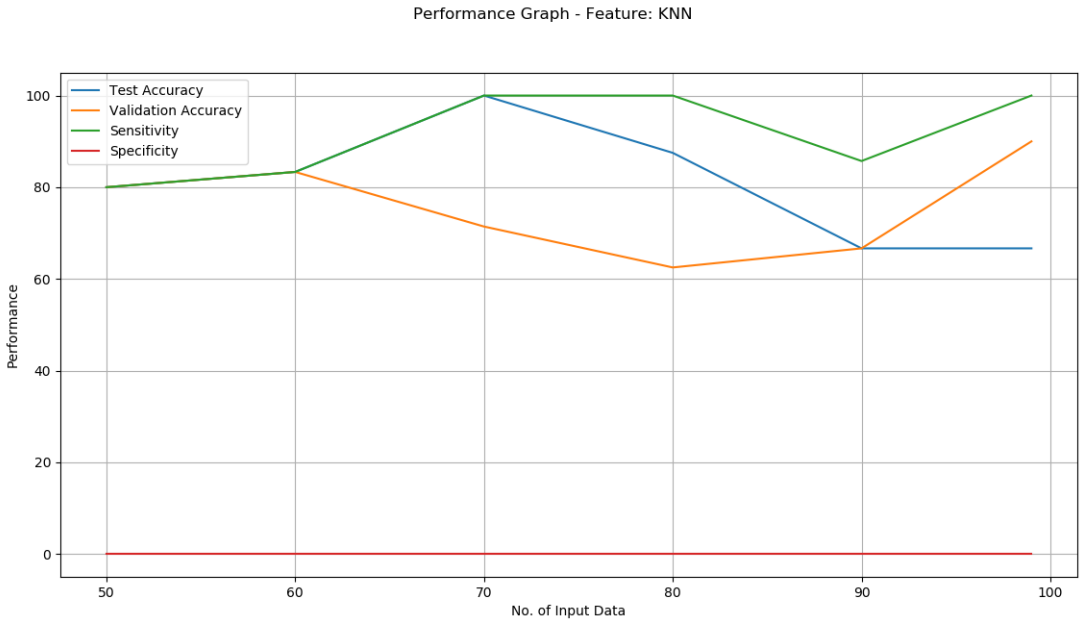

Figure 4.12: Performance Graph for Classification of simulated EMG Signal using K-Nearest Neighbor for varying length of input dataset.

using a second order butterworth filter. Then the time domain signal is converted to its corresponding time-frequency domain representation and its constituent MUAP waveform using EMG decomposition. Features are then extracted from the three domains and the best features are selected using CST and RFE. The best features obtained individually from the two feature selection techniques(FST) are then used for classification and evaluation of the signals. RFA and KNN were chosen as classifiers and their performance was evaluated using ROC Curve.

4.2.1 Feature Selection Technique

Two different FSTs were used for selection of the best features from candidate features obtained from each of the three domain representations. The results are discussed below.

Chi-Squared Test

Figure 4.13 shows the normalized scores obtained for each feature (x-axis) using CST. The scores represents the number of times each feature was selected by CST during feature selection using a 5-fold Cross Validation. The score(y-axis) of a feature is proportional to the dependence of that that feature on the given output classes. Therefore, features with higher score are more likely to be relevant for classification. Therefore, 'n' features with highest scores were selected as the best-feature-set for classification. Here, 'n' is the hyper-parameter of CST which was selected as to provide the most optimum classification performance. The optimal value of 'n' was found to be 15.
Table 4.6 shows the top 15 features selected using CST from the time-series signal, time-frequency representation of the signal and the average MUAP waveform calculated for 8 individual motor units via EMG Decomposition.

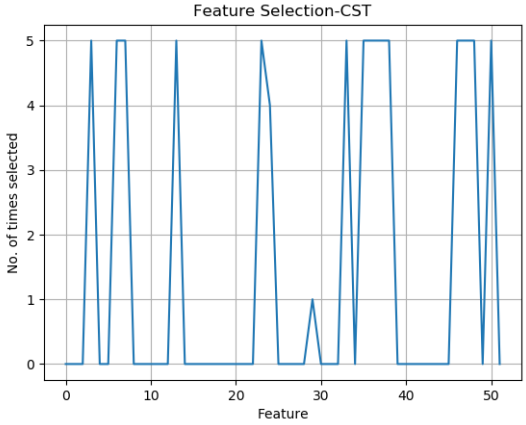

Figure 4.13: Normalized scores obtained for each feature using CST.

1. V.Dur	6. ISI	11. SSC
2. V.Amp	7. Var	12. ASA
3. A.Are	8. ZCR	13. TOP
4. V.Are	9. ZLA	14. VCF
5. V. Min and Max Amp.	10. WVL	15.VMF

Table 4.6: Best 15 features selected using CST and 5-Fold Cross Validation

Recursive Feature Elimination

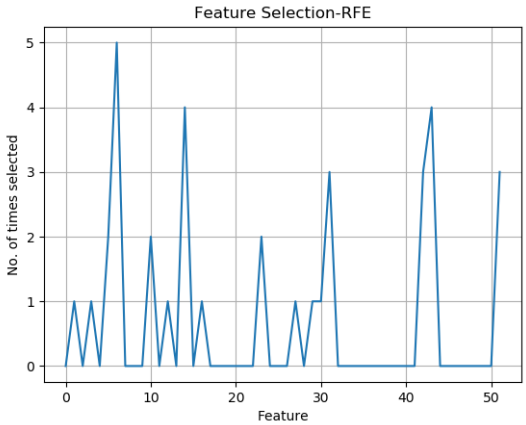

Figure 4.14: Normalized scores obtained for each feature using RFE.

Figure 4.14 shows the scores obtained for each feature (x-axis) using RFE. The scores represents the number of times each feature was selected by RFE algorithm during feature selection using a 5-fold Cross Validation. The score(y-axis) of a feature is proportional to the dependence of that that feature on the given output classes. Since RFE has a requirement of predefined regression model, Logistic Regression was used as the model for selecting the best features using RFE.
Table 4.7 shows the top 15 features selected from the time-series signal, time-frequency representation of the signal and the average MUAP waveform calculated for 8 individual motor units via EMG Decomposition. RFE and a linear regression

1. A.Amp	6. S.Min Max.Amp	11. Kur
2. A.Ris	7. A.Dur	12. MMAV
3. RMS	8. TOP	13. STD
4. VAR	9. MNF	14. AMD
5. Min.Amp	10. V.Ris	15. V. Min and Max Amp.

Table 4.7: Best 15 features selected using RFE and 5-Fold Cross Validation

model was used for feature selection.

4.2.2 Classification

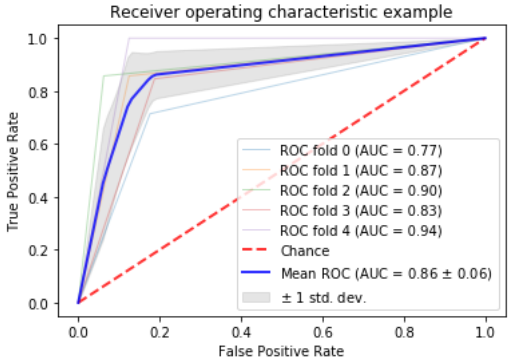

Figure 4.15: ROC Curve for classification using random forest algorithm and feature set extracted via chi-squared test.

Figure 4.15 shows the ROC curves obtained after classification of the two best-feature-sets chosen via CST. The classifications were performed using RFA and KNN. The classification was performed using 5-fold Cross Validation. The experiments were conducted for different sizes of best-feature-set and the selecting 15 best-feature-sets provided the most optimal outcome in each case. Only results of the best outcome are shown here. It can be observed from the above ROC Curve that the features selected using CST provided the best accuracy(84-86%) when the classification was performed using RFA. The hyper-parameters i.e. the number of decision trees for RFA and the number of nearest neighbors for KNN was selected using PSO.
While performing classification with 5-fold Cross Validation, the dataset was first split into two parts - one containing the complete training data and the other containing evaluation data. The training dataset was further split into two parts - one containing partial training data used for training the PSO algorithm and the other containing test data which was used to choose the best hyper-parameters for the respective classifiers. The classifiers were then trained using the complete training data and performance of the classifiers were evaluated using the evaluation data.
Table 4.8 shows the performance in terms of ROC curve calculated during 5-fold Cross Validation. The True Positive Rate(TPR), False Positive Rate(FPR) and Area under Curve(AUC) is calculated for different combinations of classifiers(CLF) and feature sets selected using different feature selection techniques(FST). According to Table 4.8, the 15 best features mentioned in Table 4.6 provides the best

FST	CLF	TPR	FPR	AUC
CST	RFA	86%	20%	85% ±2%
RFE	RFA	78%	25%	75% ±9%
CST	KNN	79%	32%	77% ±6%
RFE	KNN	78%	30%	75% ±9%

Table 4.8: Performance Table for Classification using Multi-Domain Feature Extraction

accuracy(85%) and TPR(86%) along with a low FPR(20%) compared to rest of the other techniques when the classification is performed using RFA.

4.2.3 Analysis

Accuracy(%)	SE(%)	SP(%)	Methods
74	80	60	[26])
85	**86**	**80**	**Proposed**
68	70	63	[78]
77	85	60	[90]

Table 4.9: Performance Comparison with Existing Systems

Table 4.9 shows the comparison(accuracy, sensitivity-se and specificity-sp) between the proposed method and different existing methods. The performance metric was obtained after implementing the methods proposed in each paper with our dataset. This was done in order to make a comparison of the performances of these algorithms when classification is performed with the same dataset. It can be observed that the proposed method shows a high accuracy, specificity and sensitivity even with small dataset.

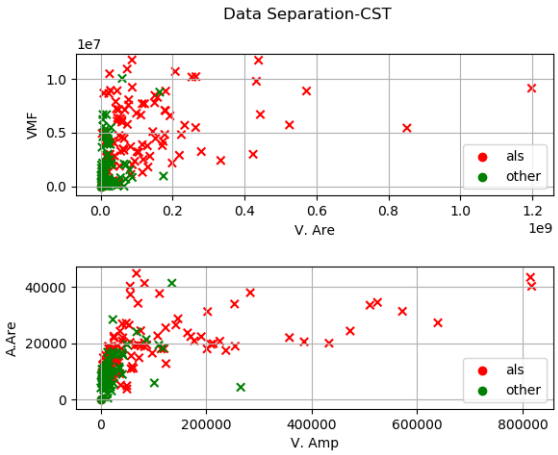

Figure 4.16: Separation of the two classification output classes based on top 4 features selected using CST

Figure 4.16 and 4.17 shows the scatter plot for separation between two output classes(ALS data and healthy data) of the classification algorithm using highest and

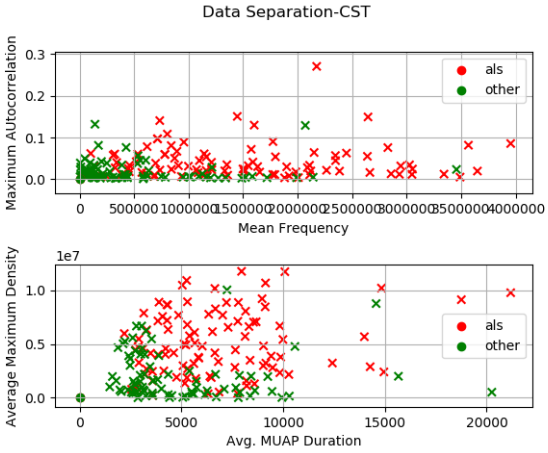

Figure 4.17: Separation of the two classification output classes based on least scoring 4 features selected using CST

least ranking 4 features respectively, evaluated using CST. Each axis of the graph represents a best-feature selected and even though the graphs of Fig. 4.16 show a pretty distinguishable separation between healthy subjects and subjects affected with ALS, it can be observed that a small fraction of ALS data overlaps with healthy data. It might be due to the fact that some of the patients included in the dataset were diagnosed within a very short duration (less than 1 year) since the onset of disease as mentioned in [41] or probably due to the fact that not all EMG signals acquired from a specific patient showed abnormal activity. Outliers were not removed from the dataset in order to preserve the context of stochasticity and adaptivity. It can further be observed that the Variance of Rectified MUAP Area(V.Are.) and variance of MUAP amplitude(V.Amp) obtained via EMG decomposition tends to be concentrated at a lower range whereas for healthy subjects whereas it tends to shift towards higher region of the feature space for ALS subjects. This might be due to the fact that due to changes in muscle fibers, the area and amplitude of MUAP waveform tends to be larger for patients suffering from ALS[14]. Similarly, Mean Frequency(MNF), Average MUAP waveform Duration(A.Dur) and Average Maximum Density(AMD) of Fig. 4.17 creates a visible separation of classes with higher overlap.

Bibliography

[1] K. Pearson, "X. on the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling", *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 50, no. 302, pp. 157–175, 1900.

[2] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity", *The bulletin of mathematical biophysics*, vol. 5, no. 4, pp. 115–133, 1943.

[3] A. L. Hodgkin and A. F. Huxley, "A quantitative description of membrane current and its application to conduction and excitation in nerve", *The Journal of physiology*, vol. 117, no. 4, pp. 500–544, 1952.

[4] F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain.", *Psychological review*, vol. 65, no. 6, p. 386, 1958.

[5] R. Burke, D. Levine, P. Tsairis, and F. 3. Zajac, "Physiological types and histochemical profiles in motor units of the cat gastrocnemius", *The Journal of physiology*, vol. 234, no. 3, pp. 723–748, 1973.

[6] R. Warwick, P. L. Williams, and H. Gray, *Gray's anatomy*. Longman, 1973.

[7] D. Burke, N. Skuse, and D. Stuart, "The regularity of muscle spindle discharge in man.", *The Journal of physiology*, vol. 291, no. 1, pp. 277–290, 1979.

[8] F. Buchthal and H. Schmalbruch, "Motor unit of mammalian muscle.", *Physiological Reviews*, vol. 60, no. 1, pp. 90–142, 1980.

[9] R. S. LeFever and C. J. De Luca, "A procedure for decomposing the myoelectric signal into its constituent action potentials-part i: Technique, theory, and implementation", *IEEE transactions on biomedical engineering*, no. 3, pp. 149–157, 1982.

[10] R. N. Bracewell and R. N. Bracewell, *The Fourier transform and its applications*. McGraw-Hill New York, 1986, vol. 31999.

[11] J. R. Quinlan, "Induction of decision trees", *Machine learning*, vol. 1, no. 1, pp. 81–106, 1986.

[12] D. S. Broomhead and D. Lowe, "Radial basis functions, multi-variable functional interpolation and adaptive networks", Royal Signals and Radar Establishment Malvern (United Kingdom), Tech. Rep., 1988.

[13] L. J. Dorfman and K. C. McGill, "Aaee minimonograph# 29: Automatic quantitative electromyography", *Muscle & Nerve: Official Journal of the American Association of Electrodiagnostic Medicine*, vol. 11, no. 8, pp. 804–818, 1988.

[14] S. D. Nandedkar, P. E. Barkhaus, D. B. Sanders, and E. V. Stålberg, "Analysis of amplitude and area of concentric needle emg motor unit action potentials", *Electroencephalography and clinical neurophysiology*, vol. 69, no. 6, pp. 561–567, 1988.

[15] W. Jänig, "Autonomic nervous system", in *Human physiology*, Springer, 1989, pp. 333–370.

[16] B. R. Brooks, R. Depaul, Y. De Tan, M. Sanjak, R. L. Sufit, and J. Robbins, "Motor neuron disease", in *Controlled clinical trials in neurological disease*, Springer, 1990, pp. 249–281.

[17] E. Kleinberg, "Stochastic discrimination", *Annals of Mathematics and Artificial intelligence*, vol. 1, no. 1, pp. 207–239, 1990.

[18] C. Schizas, C. Pattichis, I. Schofield, P. Fawcett, and L. Middleton, "Artificial neural nets in computer-aided macro motor unit potential classification", *IEEE Engineering in Medicine and Biology Magazine*, vol. 9, no. 3, pp. 31–38, 1990.

[19] N. S. Altman, "An introduction to kernel and nearest-neighbor nonparametric regression", *The American Statistician*, vol. 46, no. 3, pp. 175–185, 1992.

[20] D. R. Cornblath, R. W. Kuncl, E. D. Mellits, S. A. Quaskey, L. Clawson, A. Pestronk, and D. B. Drachman, "Nerve conduction studies in amyotrophic lateral sclerosis", *Muscle & Nerve: Official Journal of the American Association of Electrodiagnostic Medicine*, vol. 15, no. 10, pp. 1111–1115, 1992.

[21] M. H. Hassoun, C. Wang, and A. Spitzer, "Nnerve: Neural network extraction of repetitive vectors for electromyography. i. algorithm", *IEEE Transactions on Biomedical Engineering*, vol. 41, no. 11, pp. 1039–1052, 1994.

[22] P. N. Leigh and K. Ray-Chaudhuri, "Motor neuron disease.", *Journal of neurology, neurosurgery, and psychiatry*, vol. 57, no. 8, p. 886, 1994.

[23] C. Cortes and V. Vapnik, "Support-vector networks", *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[24] T. K. Ho, "Random decision forests", in *Proceedings of 3rd international conference on document analysis and recognition*, IEEE, vol. 1, 1995, pp. 278–282.

[25] H. Lodish, A. Berk, S. L. Zipursky, P. Matsudaira, D. Baltimore, J. Darnell, *et al.*, *Molecular cell biology*. WH Freeman New York, 1995, vol. 3.

[26] C. S. Pattichis, C. N. Schizas, and L. T. Middleton, "Neural network models in emg diagnosis", *IEEE transactions on biomedical Engineering*, vol. 42, no. 5, pp. 486–496, 1995.

[27] P. C. Wong, C. A. Pardo, D. R. Borchelt, M. K. Lee, N. G. Copeland, N. A. Jenkins, S. S. Sisodia, D. W. Cleveland, and D. L. Price, "An adverse property of a familial als-linked sod1 mutation causes motor neuron disease characterized by vacuolar degeneration of mitochondria", *Neuron*, vol. 14, no. 6, pp. 1105–1116, 1995.

[28] M. Zardoshti-Kermani, B. C. Wheeler, K. Badie, and R. M. Hashemi, "Emg feature evaluation for movement control of upper extremity prostheses", *IEEE Transactions on Rehabilitation Engineering*, vol. 3, no. 4, pp. 324–333, 1995.

[29] K. Ben and P. Van der Smagt, "An introduction to neural networks", *University of Amsterdam*, 1996.

[30] S. Hochreiter and J. Schmidhuber, "Long short-term memory", *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[31] A. Hof, "The relationship between electromyogram and muscle force", *Sportverletzung· Sportschaden*, vol. 11, no. 03, pp. 79–86, 1997.

[32] R. Watanabe, M. Iino, M. Honda, J. Sano, and M. Hara, "Primary lateral sclerosis", *Neuropathology*, vol. 17, no. 3, pp. 220–224, 1997.

[33] Y. Yang and J. O. Pedersen, "A comparative study on feature selection in text categorization", in *Icml*, vol. 97, 1997, p. 35.

[34] J. Kwon, S. Lee, C. Shin, Y. Jang, and S. Hong, "Signal hybrid hmm-ga-mlp classifier for continuous emg classification purpose", in *Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Vol. 20 Biomedical Engineering Towards the Year 2000 and Beyond (Cat. No. 98CH36286)*, IEEE, vol. 3, 1998, pp. 1404–1407.

[35] A. Vermeulen and J.-P. Rospars, "Dendritic integration in olfactory sensory neurons: A steady-state analysis of how the neuron structure and neuron environment influence the coding of odor intensity", *Journal of computational neuroscience*, vol. 5, no. 3, pp. 243–266, 1998.

[36] C. I. Christodoulou and C. S. Pattichis, "Unsupervised pattern recognition for the classification of emg signals", *IEEE Transactions on Biomedical Engineering*, vol. 46, no. 2, pp. 169–178, 1999.

[37] R. Merletti and P. Di Torino, "Standards for reporting emg data", *J Electromyogr Kinesiol*, vol. 9, no. 1, pp. 3–4, 1999.

[38] C. S. Pattichis and M. S. Pattichis, "Time-scale analysis of motor unit action potentials", *IEEE Transactions on Biomedical Engineering*, vol. 46, no. 11, pp. 1320–1329, 1999.

[39] J. Friedman, T. Hastie, and R. Tibshirani, *The elements of statistical learning*, 10. Springer series in statistics New York, NY, USA: 2001, vol. 1.

[40] N. Lechtzin, C. Wiener, L. Clawson, V. Chaudhry, and G. Diette, "Hospitalization in amyotrophic lateral sclerosis: Causes, costs, and outcomes", *Neurology*, vol. 56, no. 6, pp. 753–757, 2001.

[41] N. M, "Detailed analysis of clinical electromyography signals emg decomposition, findings and firing pattern analysis in controls and patients with myopathy and amytrophic lateral sclerosis", 2001. [Online]. Available: http://www.emglab.net.

[42] D. T. Mewett, H. Nazeran, and K. J. Reynolds, "Removing power line noise from recorded emg", in *Engineering in Medicine and Biology Society, 2001. Proceedings of the 23rd Annual International Conference of the IEEE*, IEEE, vol. 3, 2001, pp. 2190–2193.

[43] D. Purves, G. Augustine, D. Fitzpatrick, L. Katz, A. LaMantia, J. McNamara, and S. Williams, *Neuroscience 2nd edition. sunderland (ma) sinauer associates*, 2001.

[44] R. Boostani and M. H. Moradi, "Evaluation of the forearm emg signal features for the control of a prosthetic hand", *Physiological measurement*, vol. 24, no. 2, p. 309, 2003.

[45] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection", *Journal of machine learning research*, vol. 3, no. Mar, pp. 1157–1182, 2003.

[46] E. Cantú-Paz, S. Newsam, and C. Kamath, "Feature selection in scientific applications", in *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, 2004, pp. 788–793.

[47] D. U. Silverthorn, W. C. Ober, C. W. Garrison, A. C. Silverthorn, and B. R. Johnson, *Human physiology: an integrated approach*. Pearson/Benjamin Cummings San Francisco, CA: 2004.

[48] A. Hamilton-Wright and D. W. Stashuk, "Physiologically based simulation of clinical emg signals", *IEEE Transactions on biomedical engineering*, vol. 52, no. 2, pp. 171–183, 2005.

[49] K. C. McGill, Z. C. Lateva, and H. R. Marateb, "Emglab: An interactive emg decomposition program", *Journal of neuroscience methods*, vol. 149, no. 2, pp. 121–133, 2005.

[50] S. C. Barber, R. J. Mead, and P. J. Shaw, "Oxidative stress in als: A mechanism of neurodegeneration and a therapeutic target", *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, vol. 1762, no. 11-12, pp. 1051–1067, 2006.

[51] J. K. Fink, "Hereditary spastic paraplegia", *Current neurology and neuroscience reports*, vol. 6, no. 1, pp. 65–76, 2006.

[52] P. M. Granitto, C. Furlanello, F. Biasioli, and F. Gasperi, "Recursive feature elimination with random forest for ptr-ms analysis of agroindustrial products", *Chemometrics and Intelligent Laboratory Systems*, vol. 83, no. 2, pp. 83–90, 2006.

[53] E. W. Sellers and E. Donchin, "A p300-based brain–computer interface: Initial tests by als patients", *Clinical neurophysiology*, vol. 117, no. 3, pp. 538–548, 2006.

[54] L. K. McCorry, "Physiology of the autonomic nervous system", *American journal of pharmaceutical education*, vol. 71, no. 4, p. 78, 2007.

[55] J. D. Allen, "Human physiology-the basis of medicine", *The Ulster medical journal*, vol. 77, no. 3, p. 216, 2008.

[56] Z. Erim and W. Lin, "Decomposition of intramuscular emg signals using a heuristic fuzzy expert system", *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 9, pp. 2180–2189, 2008.

[57] S. Love, D. Louis, and D. W. Ellison, *Greenfield's Neuropathology Eighth Edition 2-Volume Set*. CRC Press, 2008.

[58] G. J. Tortora and B. H. Derrickson, *Principles of anatomy and physiology*. John Wiley & Sons, 2008.

[59] B. Pandey and R. Mishra, "An integrated intelligent computing model for the interpretation of emg based neuromuscular diseases", *Expert Systems with Applications*, vol. 36, no. 5, pp. 9201–9213, 2009.

[60] A. Shmilovici, "Support vector machines", in *Data mining and knowledge discovery handbook*, Springer, 2009, pp. 231–247.

[61] M. R. Canal, "Comparison of wavelet and short time fourier transform methods in the analysis of emg signals", *Journal of medical systems*, vol. 34, no. 1, pp. 91–94, 2010.

[62] N. S. Full, "Decision tree algorithm", 2010.

[63] P. A. Kaplanis, C. S. Pattichis, D. Zazula, *et al.*, "Multiscale entropy-based approach to automated surface emg classification of neuromuscular disorders", *Medical & biological engineering & computing*, vol. 48, no. 8, pp. 773–781, 2010.

[64] A. Subasi and M. K. Kiymik, "Muscle fatigue detection in emg using time–frequency methods, ica and neural networks", *Journal of medical systems*, vol. 34, no. 4, pp. 777–785, 2010.

[65] D. Tkach, H. Huang, and T. A. Kuiken, "Study of stability of time-domain features for electromyographic pattern recognition", *Journal of neuroengineering and rehabilitation*, vol. 7, no. 1, p. 21, 2010.

[66] Y. Wu, S. Krishnan, and R. M. Rangayyan, "Computer-aided diagnosis of knee-joint disorders via vibroarthrographic signal analysis: A review", *Critical Reviews^{TM} in Biomedical Engineering*, vol. 38, no. 2, 2010.

[67] Y. Wu and S. C. Ng, "A pdf-based classification of gait cadence patterns in patients with amyotrophic lateral sclerosis", in *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, IEEE, 2010, pp. 1304–1307.

[68] A. F. Bastos, M. Orsini, D. Machado, M. P. Mello, S. Nader, J. G. Silva, A. M. da Silva Catharino, M. R. de Freitas, A. Pereira, L. L. Pessoa, *et al.*, "Amyotrophic lateral sclerosis: One or multiple causes?", *Neurology international*, vol. 3, no. 1, 2011.

[69] W. A. N. Dorland, *Dorland's Illustrated Medical Dictionary32: Dorland's Illustrated Medical Dictionary*. Elsevier Health Sciences, 2011.

[70] M. C. Kiernan, S. Vucic, B. C. Cheah, M. R. Turner, A. Eisen, O. Hardiman, J. R. Burrell, and M. C. Zoing, "Amyotrophic lateral sclerosis", *The Lancet*, vol. 377, no. 9769, pp. 942–955, 2011.

[71] A. Phinyomark, S. Hirunviriya, A. Nuidod, P. Phukpattaranont, and C. Limsakul, "Evaluation of emg feature extraction for movement control of upper limb prostheses based on class separation index", in *5th Kuala Lumpur International Conference on Biomedical Engineering 2011*, Springer, 2011, pp. 750–754.

[72] A. Phinyomark, P. Phukpattaranont, and C. Limsakul, "Wavelet-based denoising algorithm for robust emg pattern recognition", *Fluctuation and Noise Letters*, vol. 10, no. 02, pp. 157–167, 2011.

[73] J. Rafiee, M. Rafiee, F. Yavari, and M. Schoen, "Feature extraction of forearm emg signals for prosthetics", *Expert Systems with Applications*, vol. 38, no. 4, pp. 4058–4067, 2011.

[74] D. Schacter, *Psychology Second Edition*. Worth Publishers, 2011.

[75] N. Sobahi, "Denoising of emg signals based on wavelet transform", *Asian Transactions on Engineering*, vol. 1, no. 5, pp. 17–23, 2011.

[76] N. R. Carlson, *Physiology of behavior*. Pearson Higher Ed, 2012.

[77] A. S. U. Doulah, M. A. Iqbal, and M. A. Jumana, "Als disease detection in emg using time-frequency method", in *2012 International Conference on Informatics, Electronics & Vision (ICIEV)*, IEEE, 2012, pp. 648–651.

[78] S. A. Fattah, M. A. Iqbal, M. A. Jumana, and A. S. U. Doulah, "Identifying the motor neuron disease in emg signal using time and frequency domain features with comparison", *Signal & Image Processing*, vol. 3, no. 2, p. 99, 2012.

[79] N. K. Gosall and G. S. Gosall, *The doctor's guide to critical appraisal*. PasTest Ltd, 2012.

[80] D. C. Montgomery, E. A. Peck, and G. G. Vining, *Introduction to linear regression analysis*. John Wiley & Sons, 2012, vol. 821.

[81] A. Phinyomark, P. Phukpattaranont, and C. Limsakul, "Feature reduction and selection for emg signal classification", *Expert systems with applications*, vol. 39, no. 8, pp. 7420–7431, 2012.

[82] P. Pongpanitanont and W. Charoensuk, "Emg classification using the second order volterra series", in *The 5th 2012 Biomedical Engineering International Conference*, IEEE, 2012, pp. 1–3.

[83] L. Squire, D. Berg, F. E. Bloom, S. Du Lac, A. Ghosh, and N. C. Spitzer, *Fundamental neuroscience*. Academic Press, 2012.

[84] A. Subasi, "Classification of emg signals using combined features and soft computing techniques", *Applied soft computing*, vol. 12, no. 8, pp. 2188–2198, 2012.

[85] D. Laight, "Overview of peripheral nervous system pharmacology", *Nurse Prescribing*, vol. 11, no. 9, pp. 448–454, 2013.

[86] R. S. Michalski, J. G. Carbonell, and T. M. Mitchell, *Machine learning: An artificial intelligence approach*. Springer Science & Business Media, 2013.

[87] A. Phinyomark, F. Quaine, S. Charbonnier, C. Serviere, F. Tarpin-Bernard, and Y. Laurillau, "Emg feature evaluation for improving myoelectric pattern recognition robustness", *Expert Systems with applications*, vol. 40, no. 12, pp. 4832–4840, 2013.

[88] G. Pocock, C. D. Richards, D. Richards, and D. A. Richards, *Human physiology*. Oxford university press, 2013.

[89] G. Robertson, G. Caldwell, J. Hamill, G. Kamen, and S. Whittlesey, *Research methods in biomechanics, 2E*. Human Kinetics, 2013.

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

[91] T. W. Beck, M. S. Stock, and J. M. Defreitas, "Shifts in emg spectral power during fatiguing dynamic contractions", *Muscle & nerve*, vol. 50, no. 1, pp. 95–102, 2014.

[92] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using rnn encoder-decoder for statistical machine translation", *arXiv preprint arXiv:1406.1078*, 2014.

[93] K. Rajdl and P. Lansky, "Fano factor estimation", *Math Biosci Eng*, vol. 11, pp. 105–123, 2014.

[94] M. Zhu, W. Xiong, and Y.-F. B. Wu, "Learning to rank with only positive examples", in *2014 13th International Conference on Machine Learning and Applications*, IEEE, 2014, pp. 87–92.

[95] T. Liewluck and D. S. Saperstein, "Progressive muscular atrophy", *Neurologic clinics*, vol. 33, no. 4, pp. 761–773, 2015.

[96] santini. (2015). Decision trees (2), [Online]. Available: https://www.slideshare.net/marinasantini1/lecture-4-decision-trees-2-entropy-information-gain-gain-ratio-55241087.

[97] C. Altın and O. Er, "Comparison of different time and frequency domain feature extraction methods on elbow gesture's emg", *European journal of interdisciplinary studies*, vol. 2, no. 3, pp. 35–44, 2016.

[98] Brownlee. (2016). Supervised and unsupervised machine learning algorithms, [Online]. Available: https://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/.

[99] M. O. Krucoff, S. Rahimpour, M. W. Slutzky, V. R. Edgerton, and D. A. Turner, "Enhancing nervous system recovery through neurobiologics, neural interface training, and neurorehabilitation", *Frontiers in neuroscience*, vol. 10, p. 584, 2016.

[100] MOSELEY. (2016). Explainer – what is pain?, [Online]. Available: https://bodyinmind.org/what-is-pain/.

[101] G. R. Naik, S. E. Selvan, and H. T. Nguyen, "Single-channel emg classification with ensemble-empirical-mode-decomposition-based ica for diagnosing neuromuscular disorders", *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 24, no. 7, pp. 734–743, 2016.

[102] J. P. Taylor, R. H. Brown Jr, and D. W. Cleveland, "Decoding als: From genes to mechanism", *Nature*, vol. 539, no. 7628, p. 197, 2016.

[103] A. V. H. Amrutha N, "A review on noises in emg signal and its removal", *International Journal of Scientific and Research Publications*, vol. 7, May 2017, ISSN: 2250-3153. [Online]. Available: http://www.ijsrp.org/research-paper-0517/ijsrp-p6504.pdf.

[104] Y. Martınez-Campo, C. Homedes, A. Lazaro, R. Alarcón, D. Campo, M. Riera, R. Domınguez, M. Povedano, and C. Casasnovas, "Observational study of patients in spain with amyotrophic lateral sclerosis: Correlations between clinical status, quality of life, and dignity", *BMC palliative care*, vol. 16, no. 1, p. 75, 2017.

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

[106] N. Sengar, M. K. Dutta, and C. M. Travieso, "Identification of amyotrophic lateral sclerosis using emg signals", in *2017 4th IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics (UPCON)*, IEEE, 2017, pp. 468–471.

[107] S. Angelova, S. Ribagin, R. Raikova, and I. Veneva, "Power frequency spectrum analysis of surface emg signals of upper limb muscles during elbow flexion–a comparison between healthy subjects and stroke survivors", *Journal of Electromyography and Kinesiology*, vol. 38, pp. 7–16, 2018.

[108] Donges. (). The random forest algorithm, [Online]. Available: https://towardsdatascience.com/the-random-forest-algorithm-d457d499ffcd.

[109] R. Drake, W. Vogl, and A. Mitchell, "Grey's anatomy for students. philadelphia, 2004", ISBN 0-443-06612-4, Tech. Rep.

[110] Knierim. (). Chapter 1: Motor units and muscle receptors, [Online]. Available: https://nba.uth.tmc.edu/neuroscience/s3/chapter01.html.